Auditory localization under conflicting dynamic and spectral cues

Henri Pöntynen

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 31.07.2015

Thesis supervisor:

Prof. Ville Pulkki

Thesis advisor:

M.Sc. (Tech.) Olli Santala

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Henri Pöntynen

Title: Auditory localization under conflicting dynamic and spectral cues

Date: 31.07.2015 Language: English Number of pages: 10+60

Department of Signal Processing and Acoustics

Professorship: Acoustics and Audio Signal Processing Code: S-89

Supervisor: Prof. Ville Pulkki

Advisor: M.Sc. (Tech.) Olli Santala

Sound localization in humans functions by identifying and comparing different attributes in acoustic signals presented to the two ears. These attributes are largely independent of one another and can thus be manipulated separately to provide the auditory system with conflicting cues to sound source location. While the topic of cue conflict between binaural cues has received a fair amount of attention in the field of psychoacoustics, localization of acoustic stimuli providing discrepant dynamic and spectral cues remains a less explored topic.

The experimental part of this thesis consists of the design, implementation and documentation of a psychoacoustic experiment assessing cue conflict. In the implemented listening test, subjects performed horizontal head rotations while being exposed to noise burst sequences presented over loudspeakers in the horizontal plane. Head orientation data gathered via motion tracking cameras was used to move the stimuli in tandem with the subject's head movements to artificially alter the rate of change in the binaural cues during head movements. The resulting binaural cues corresponded to those produced by an overhead source. Subjects were asked to report whether they localized the stimuli to the horizontal direction indicated by the spectral cues or not.

The results from the conducted experiment indicate wide intersubject variability in localization performance. In general, subjects exposed to stimuli providing conflicting dynamic and spectral cues either display a preference for one of the two cues or switch the cue they attend to between trials, leading to inconsistent localization performance. Results from some subjects displayed an apparent localization bias towards the rear hemiplane.

Keywords: Psychoacoustics, dynamic cues, spectral cues, cue conflict

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Henri Pöntynen

Työn nimi: Äänen lokalisaatio ristiriitaisten spektraalisten ja dynaamisten lokalisaatiovihjeiden perusteella

Päivämäärä: 31.07.2015 Kieli: Englanti Sivumäärä: 10+60

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Akustiikka ja äänenkäsittelytekniikka Koodi: S-89

Työn valvoja: Prof. Ville Pulkki

Työn ohjaaja: DI Olli Santala

Ihmisen suuntakuulo toimii tunnistamalla ja vertailemalla piirteitä korviin saapuvista akustisista signaaleista. Suuntakuulon kannalta tärkeät signaaliparametrit ovat pitkälti toisistaan riippumattomia, jolloin herätteen kantamat lokalisaatiovihjeet voidaan asettaa toisiinsa nähden ristiriitaisiksi. Siinä missä ristiriitaisten binauraalisten lokalisaatiovihjeiden muodostamia lokalisaatiohavaintoja on tutkittu laajalti, dynaamisten ja spektraalisten lokalisaatiovihjeiden vastaava tutkimus on ollut vähäistä.

Opinnäytetyön kokeellisessa osassa suunniteltiin, toteutettiin ja dokumentoitiin psykoakustinen koe, jossa tutkittiin ristiriitaisien lokalisaatiovihjeiden vaikutusta äänen lokalisaatioon. Toteutetussa kokeessa koehenkilöille soitettiin kohinapurskesekvenssejä horisontaalitasossa sijaitsevista kaiuttimista pään liikkeiden aikana. Koehenkilöiden pään asentoa seurattiin kamerajärjestelmällä, joka mahdollisti herätteiden esityssuunnan vaihtamisen pään kulloisenkin asennon mukaisesti. Kohinapurskeiden esityssuuntaa muutettiin siten, että korvien väliset signaalierot eivät muuttuneet pään liikkeiden aikana, jolloin dynaamiset lokalisaatiovihjeet viittasivat pään yläpuolella olevaan äänilähteeseen. Koehenkilöiden tehtävänä oli ilmoittaa lokalisoituivatko esitetyt purskesekvenssit spektrivihjeiden mukaiseen suuntaan horisontaalitasossa.

Kokeen tulokset osoittavat lokalisaatiotarkkuuden vaihtelevan laajasti koehenkilöiden välillä. Pieni osa koehenkilöistä suosi johdonmukaisesti joko dynaamisia- tai spektrivihjeitä lokalisoidessaan esitettyjä purskeita. Useat koehenkilöistä vaihtoivat lokalisaatiohavainnon muodostuksessa suosimaansa lokalisaatiovihjettä kokeen edetessä, jolloin lokalisaatiotarkkuus oli vaihtelevaa ja epäjohdonmukaista. Joidenkin koehenkilöiden koetulokset viittaavat siihen, että ristiriitaiset vihjeet aiheuttivat takasuuntaan kohdistuvan vinouman lokalisaatiossa.

Avainsanat: Psykoakustiikka, dynaamiset lokalisaatiovihjeet, spektrivihjeet, ristiriitaiset lokalisaatiovihjeet

Acknowledgements

This work was carried out in the Department of Signal Processing and Acoustics at Aalto University School of Electrical Engineering during the spring of 2015.

First and foremost I want to thank my instructor M.Sc. Olli Santala and supervisor Prof. Ville Pulkki for their guidance, support and helpful discussions throughout the thesis work. Additionally, I'd like to express my gratitude to Prof. Pulkki for providing me the opportunity to work on this interesting topic. I also want to thank all of my co-workers at the Department of Signal Processing and Acoustics for creating such an enjoyable and supportive work environment.

Finally, I'd like to thank my family and friends for the invaluable support they have given me throughout my studies.

Helsinki, 31.07.2015

Henri Pöntynen

Contents

Symbols and abbreviations

Symbols

b_i	Description of auditory event
c	Speed of sound
$d_{j,i}$	Absolute deviation of observation j, i from treatment group median
df_b	Degrees of freedom between treatment groups
df_w	Degrees of freedom within treatment group
F	F-ratio
F_α	F-distribution cut-off score
f	Frequency
H_a	Alternative hypothesis
H_0	Null hypothesis
h_i	Auditory event
k	Number of treatment groups
MS_b	Mean squares between groups
MS_w	Mean squares within groups
N	Total number of observations
nk	Number of observations in treatment group k
p	Significance value
p_0	Threshold of hearing
r	Radius
SS_b	Sum of squares between groups
SS_w	Sum of squares within group
s_i	Sound event
T_α	Threshold value for treatment groups of equal size
$T_{\alpha,i,j}$	Threshold value for treatment groups of unequal size
$y_{k,nk}$	Observation nk in treatment group k
$\overline{y}_k$	Mean value of treatment group k
q_α	Studentized range statistic
$\tilde{y_k}$	Median value of treatment group k
γ	Lateral angle
$\Delta\alpha$	Angular displacement of sound source
$\Delta\beta$	Angular displacement of interaural axis
θ	Azimuth angle
λ	Wavelength
μ_k	Mean value of treatment group k
$\overline{\mu}$	Global mean of data set
σ_k^2	Variance in treatment group k
ϕ	Elevation angle

Abbreviations

ANOVA	Analysis of variance
ERB	Equivalent rectangular bandwidth
HRTF	Head-related transfer function
ILD	Interaural level difference
ITD	Interaural time difference
JND	Just noticeable difference
ITD_{max}	Maximum naturally occurring interaural time difference
RMS	Root-mean-square
SPL	Sound pressure level

List of Figures

List of Tables

1 Introduction

Auditory localization refers to the set of processes through which percepts of the apparent locations of sound sources in three-dimensional space are formed. Extensive research on the topic has resulted in the identification of a number of sensory mechanisms through which the auditory system localizes sound sources. Many of the physical parameters that contribute to localization are largely independent of one another and can thus be manipulated independently under laboratory conditions. Such manipulation allows the experimenter to synthesize acoustic stimuli that provide the auditory system with conflicting cues to sound source location.

While cue conflict between binaural cues has received a fair amount of attention in hearing research, only a small number of studies have assessed cue conflict between dynamic and spectral cues. Further exploration of the interaction of dynamic cues with other localization processes could prove to be a fruitful ground for future research with direct practical applications in the growing field of virtual reality.

The thesis aims to chart the current state of research on the topic of auditory localization of acoustic stimuli providing conflicting dynamic and spectral cues. The experimental part of the work explores the effect of stimulus spectrum on localization performance in an experiment where dynamic cues provide misleading cues to sound source location. The cue conflict is implemented by moving acoustic stimuli of various spectra in a horizontal loudspeaker array in tandem with head movements to artificially alter the rate of change in the binaural cues brought about by head movements.

1.1 Outline of the thesis

This thesis begins by introducing the basic properties of the human auditory system and outlining the functions performed by the auditory periphery. Chapter 3 provides an introduction to the methodology of psychoacoustic experimentation by presenting commonly employed experimental procedures and statistical tools for data analysis. The rest of the chapter gives a short overview of the phenomenon of masking, the concept of auditory filters and a related auditory scale.

Chapter 4 is devoted to the wide topic of auditory localization. The chapter begins by introducing the standard coordinate system and terminology used in sound localization studies. The main processes human observers rely on in forming localization judgements under free-field conditions are then presented along with a discussion of their limitations. The focus of the chapter is on dynamic cues and associated perceptual phenomena.

The other half of the thesis begins with an overview of existing literature on cue conflict between dynamic and spectral cues in Chapter 5. The experimental part of the thesis work is described in Chapter 6, where the conducted listening test is presented in detail. The chapter ends with a statistical analysis of the experimental results. Chapter 7 provides a discussion on the experimental results and the limitations of the employed experimental paradigm. The thesis ends in Chapter 8 where the work is summarized and proposals for the direction of future work are given.

2 The human auditory system

Generally speaking, the human auditory pathway can be divided into two sections: peripheral hearing and the auditory brain. The peripheral section consists of the organs of hearing and it functions essentially as a transducer between the physical sound pressure variations of the external world and the electrical signals of the human nervous system. The auditory brain processes the neural stimuli that eventually result in auditory perception. As the vast topic of neural processing performed by the auditory brain is beyond the scope of this work, the focus of this chapter is on providing an introducing to the fundamental properties of human hearing and the anatomy of the auditory periphery. For an introduction to the auditory brain, the interested reader can consult the appropriate chapters in [37] and [38].

2.1 Basic properties of human hearing

The dynamic range of the human auditory system is wide. The threshold of hearing p_0 is defined as the smallest sound pressure that the ear can detect. In the case of a healthy ear, the dynamic range between the threshold of hearing and the threshold of pain spans pressures between 20 μPa and about 65 Pa [37]. In order to account for such a wide dynamic range, it is convenient to use decibel units referenced to the threshold of hearing p_0. Decibel values defined in this way are referred to as sound pressure levels (SPL). Since sound pressure is a field quantity whose squared value is proportional to acoustic power, the SPL of a given pressure p expressed in Pascals can be calculated as shown below.

$$\text{SPL} = 10 \log_{10}\left(\frac{p^2}{p_0^2}\right) = 20 \log_{10}\left(\frac{p}{p_0}\right) \tag{1}$$

Similarly, the human ear responds to a wide range of frequencies. In the case of a healthy ear, this frequency range is approximately 20 Hz - 20 kHz as shown in the threshold curves of Fig. 1; the upper limit of this frequency range tends to decrease with age and long term exposure to noise. The sensitivity of the ear is at its highest in the frequency region between 2 kHz - 5 kHz and decreases towards either extreme of the audio band.

An important feature of the human auditory system is its ability to decompose an auditory stimulus into separate frequency bands; this remarkable feature of human hearing is sometimes referred to as Ohm's acoustic law [45]. It enables the auditory system to focus on a frequency band of interest in the presence of background noise. The frequency selectivity of human hearing is discussed in more detail in sections 2.2.2 and 3.5 where the functionality of the cochlea and the concept of auditory filters are discussed.

2.2 Peripheral hearing

Peripheral hearing refers to the set of organs depicted in Fig. 2, that together form the interface between the acoustical stimuli of the physical world and the human

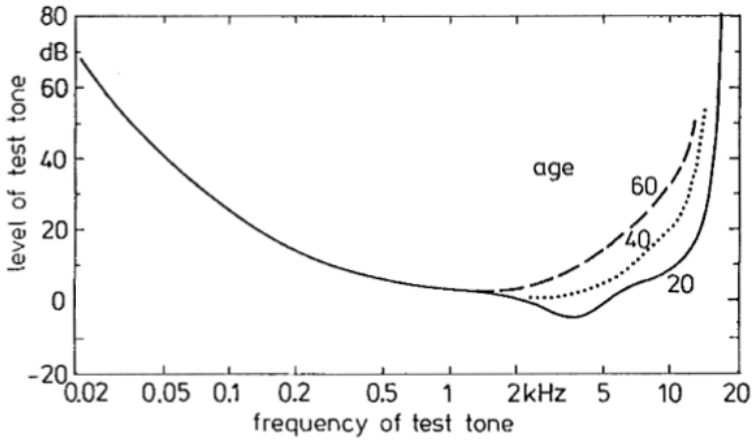


Figure 1: Typical thresholds of hearing for different age groups across the auditory band. Adopted from [8].

sensory system. Conventionally, peripheral hearing has been divided into the outer-, middle- and inner ears as shown in the schematic illustration of Fig. 3. The functions performed by the three sections of peripheral organs of hearing are described in the sections below.

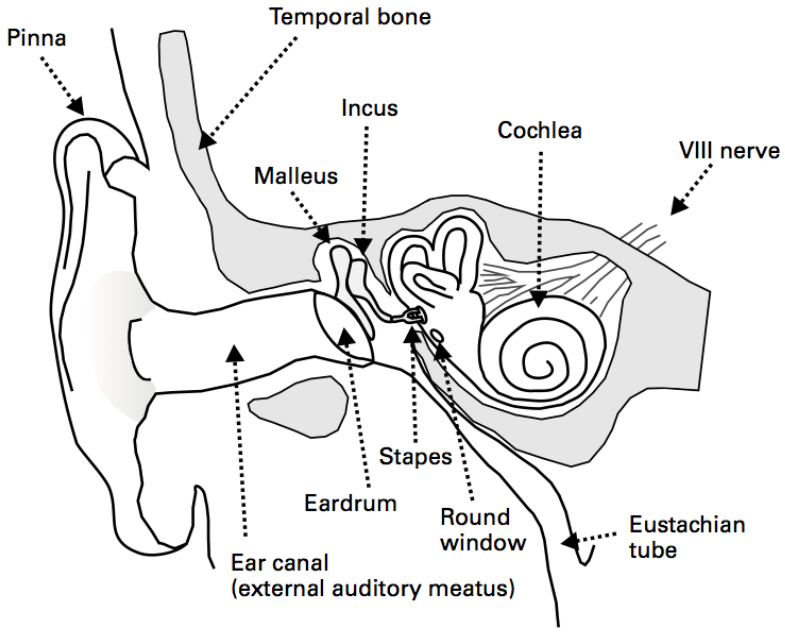


Figure 2: Cross-section of the peripheral part of the human auditory system. Adopted from [38].

2.2.1 Outer and middle ear

The outer ear consists of the pinna and the ear canal. The pinna performs two important functions. It aids in gathering acoustic energy impinging on the head and directing it towards the ear canal allowing more acoustic energy to be transferred to

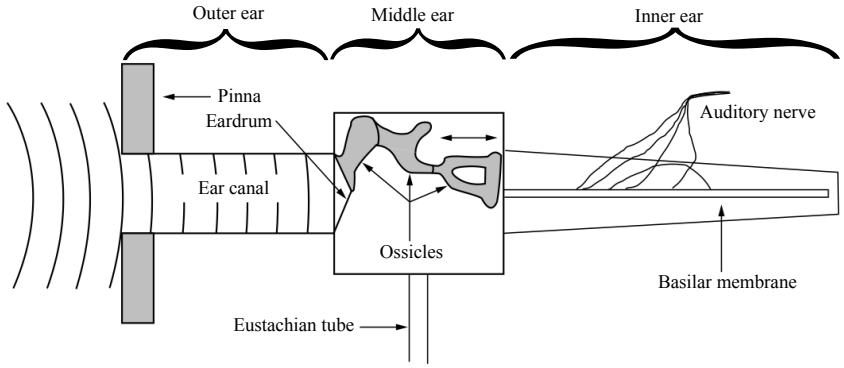


Figure 3: A simplified representation of peripheral hearing. Adapted from [36].

the eardrum. Secondly, the complex direction-dependent reflection patterns formed in the cavities of the pinna provide important localization cues for the auditory system; these cues are described in more detail in Section 4.3. The ear canal forms a resonant cavity with an approximate peak resonance of 10 dB in the frequency region of 3 - 4 kHz, thus increasing the ear's sensitivity to frequencies in this range [36]. The eardrum terminates the ear canal and joins the outer ear to the middle ear.

The middle ear consists of an air filled chamber called the tympanic cavity that houses three interconnected bones called the ossicles. The ossicular chain consist of the smallest bones in the human body: malleus, incus and stapes. The anatomy of the ossicular chain is shown on the left hand side of Fig. 4. The vibrations of the eardrum are coupled via the malleus and incus to the stapes that connects the middle ear to the inner ear through the oval window of the cochlea. The bones in the middle ear act essentially as a mechanical impedance converter between the low acoustic impedance of the air contained in the outer ear and the high impedance of the liquid contained within the inner ear [37]. An impedance mismatch causes a fraction of incident wave to be reflected back at the interface of the impedance discontinuity. By matching the impedances of the outer- and inner ears, the magnitude of this reflection is minimized and the transfer of acoustic energy to the inner ear becomes more efficient. The impedance matching is performed passively by transferring the force acting on the relatively large area of the eardrum to the small surface area of the oval window through the lever mechanism former by the ossicular chain as shown in the middle and right hand side figure of Fig. 4. Typically a human middle ear provides a fifteenfold increase in impedance between the eardrum and the oval window [37].

The tympanic cavity is connected to the oral cavity through the Eustachian tube. This connection serves as a valve that stabilizes the pressure between the tympanic cavity and the static air pressure outside of the surrounding environment.

2.2.2 Inner ear

The transduction of acoustic signals from mechanical vibrations to electrical signals in the nervous system is performed within the inner ear. The inner ear consists of the semicircular canals and the cochlea. Whereas the semicircular canals are related

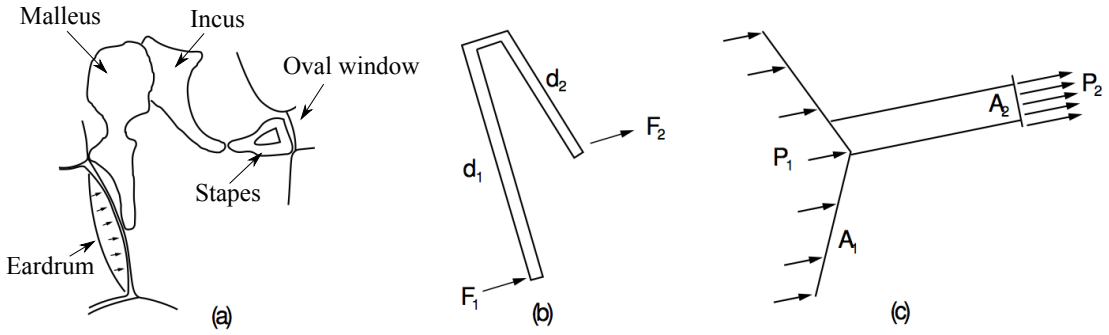


Figure 4: Ossicles within the tympanic cavity. The figure on the left hand side shows a schematic representation of the ossicular chain. The figure in the middle shows a simplified representation of the lever function performed by the ossicular chain. The figure on the right hand side illustrates the impedance transform resulting from the unequal surface areas between the eardrum and the oval window. Adapted from [17].

to the vestibular system and contribute to the formation of auditory sensations only indirectly, the cochlea is the most important organ of peripheral hearing. The cochlea is a coiled tube consisting of three longitudinal, liquid-filled chambers as shown in the cross-section on the right hand side of Fig. 5. The vibrations of the ossicular chain are coupled to the inner ear through the stapes and the oval window. The vibrating stapes sets the liquid in the cochlea in motion, which in turn moves the basilar membrane. The mechanical vibrations of the basilar membrane are then converted to eletrical signals by the hair cells on the organ of Corti, mounted on the inner rim of the basilar membrane. Referring to the grayed area in the cross-section of Fig. 5, it can be seen that the cilia—i.e., the tips—of the hair cells are in contact with the tectorial membrane, thus causing them to bend against it in proportion to the instantaneous displacement of the basilar membrane. The hair cells respond to the bending of the cilia by stimulating the auditory nerve. The mechanical vibrations of the basilar membrane are thus transduced to neural signals by the hair cells on the outer rim of the membrane.

The mechanical properties of the basilar membrane vary along its length. As illustrated by the figure on the left hand side of Fig. 5, the basilar membrane is stiff and narrow at the basal end but grows wider and more flexible towards the apical end. Due to the inhomogeneity of the membrane's mechanical properties, the width and location of the displacement maximum is a function of stimulus frequency. The wide and flexible apical end of the membrane resonates when the cochlea is stimulated with low frequencies and conversely, the basal end resonates with high-frequency stimuli; middle frequencies form resonances at intermediate distances on the membrane. The varying mechanical properties of the basilar membrane therefore allow the cochlea to function as a mechanical frequency analyzer as outlined in the figure on the right hand side of Fig. 6. The path and associated transduction processes undergone by an acoustic signal within the auditory periphery are summarized in the schematic illustration on the left hand side of Fig. 6.

A wideband stimulus can cause a number or resonances to form on the membrane simultaneously. Each resonance stimulates its associated set of hair cells that together

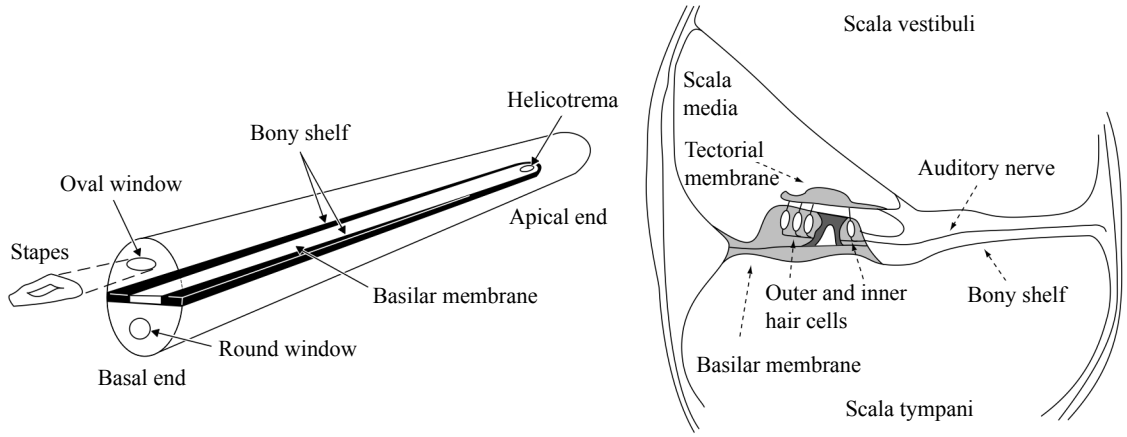


Figure 5: Simplified illustrations of the cochlea. The figure on the left hand side represents an uncoiled cochlea. The figure on the right hand shows a cross-section of the cochlea. Adapted from [17].

produce the auditory stimulus. The width of the basilar membrane resonances are closely related to the auditory system's ability to resolve individual frequency components of wide band stimuli. This is discussed in more detail in Section 3.5 on auditory filters.

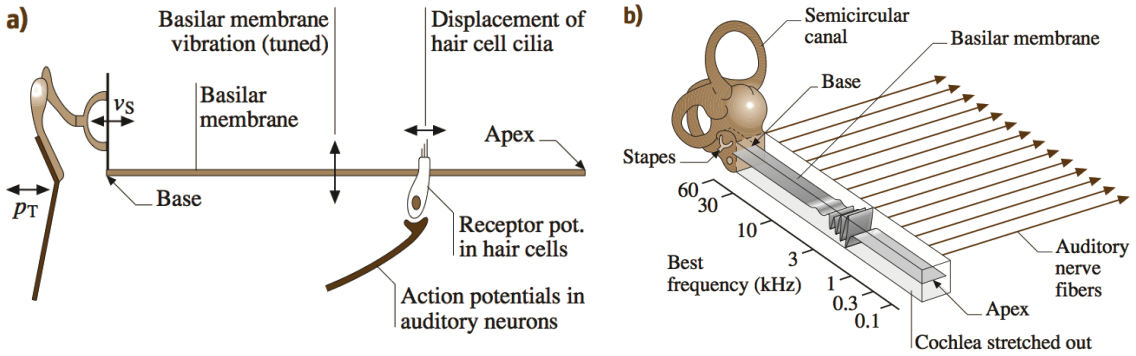


Figure 6: Schematic representation of the transduction of acoustic signals to neural activity in the auditory periphery. Figure a) shows a sideview of the uncoiled basilar membrane. The vertical movement of the basilar membrane causes a horizontal displacement in the cilia of the hair cells. The stimulated haircell activates its associated auditory neuron. Figure b) shows the frequency selectivity of the basilar membrane. Low-frequency stimuli resonate at the apex of the membrane; conversely, high frequencies resonate at the base. A set of auditory nerve fibers is connected along the length of the membrane that allows multiple signals to be sent in parallel when the cochlea is excited with wideband stimuli. Adapted from [37].

3 Psychoacoustics

The study of sensory mechanisms can be divided into physiology and psychophysics. Whereas physiological research studies sensory responses produced by physical stimuli through direct measurements from, e.g., a set of nerve fibers, psychophysical methods rely on information derived from perceptual events formed in response to physical stimuli. Psychoacoustics is a field of research that connects acoustics with psychophysics. The subject of its study is the relationship between acoustic stimuli and auditory perception. This chapter provides an introduction to selected central concepts in psychoacoustics.

3.1 Psychoacoustic measurements

Psychoacoustic experiments study the human auditory system indirectly via subjective descriptions of auditory events formed as a response to physical stimuli. The relevant factors involved in a psychoacoustic test scenario can be expressed in the simplified block diagram shown in Fig. 7. Referring to the diagram, a physical sound event s_i is mapped to a corresponding auditory event h_i via a psychophysical function $f(s)$. The psychophysical function represents the combined processing performed in the auditory periphery and the auditory brain. The experimental scenario includes a description function $f(h)$ that transfers information about the auditory event outside the cognitive space of the subject for objective analysis through statistical methods.

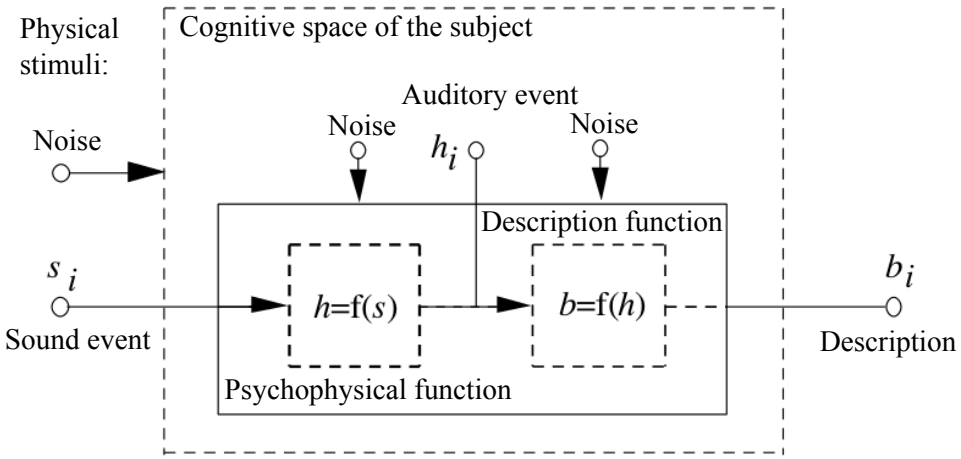


Figure 7: Block diagram representing information flow in a psychoacoustic measurement. Adapted from [17].

Because auditory events are directly accessible only to the test subject through introspection, it is clear that the description function can never provide fully accurate information; its proper choice is therefore a crucial aspect of experimental design in psychoacoustics. While there are several ways of implementing the description function, the so called forced choice procedures are probably the most common ones. In an experiment employing a forced choice answering scheme, the subject provides

a description of some aspect of an auditory event by selecting from a predetermined set of responses selected by the experimenter. As a simple example of this, a subject might be asked to describe the location of a sound source by selecting from a set of two alternatives corresponding to locations to the left and to the right of the subject. Such an arrangement limits the range of possible answers to two and is hence referred to as two-alternative forced choice.

3.1.1 Method of constant stimuli

In an experiment employing the method of constant stimuli, the experimenter selects a range of parameter values for the signal attribute under study and presents the stimuli to the test subjects in random order, independently of the performance of the subject in the chosen psychophysical task [11]. The method of constant stimuli ensures that all subjects who complete the experiment successfully are presented with the same stimuli. Results from individual subjects can then be pooled to represent the averaged effect of the signal attribute under study on the performance in the psychophysical task of the experiment.

The main drawback of the method is its inefficiency. In order for the procedure to yield reliable data, each signal attribute value needs to be presented multiple times to each subject. Moreover, the identification of the signal attribute range of interest usually requires a separate experiment to be conducted prior to the final experiment [11].

3.1.2 Adaptive staircase procedures

In an adaptive staircase procedure, the stimuli presented to the subject in a given trial are determined by the responses given to previous trials. The difficulty of the psychophysical task is adjusted from trial to trial according to the subject's responses by changing the magnitude of the signal attribute under study. If a trial results in the subject giving an incorrect answer, the experimenter adjusts the signal attribute in a manner that facilitates the psychophysical task. Conversely, a sufficient number of consequent correct answers results in the signal attribute being adjusted in a manner that makes the task more difficult. The procedure therefore adapts the difficulty of consequent trials so that the value of the modified signal attribute converges towards a threshold between consistently correct responses and responses given at chance rate. The experiment is terminated once a sufficient number of reversals has been observed; Levitt [21] recommends a minimum of six reversals until terminating the experiment. A numerical value for the threshold can be computed from the mean value of a chosen number of reversals in the direction of adjustment of the tested signal attribute. Fig. 8 shows a generic example of the results obtained from an experiment employing an adaptive procedure.

Adaptive procedures can be customized to the task at hand by adjusting the step size of the signal attribute change between trials. If an estimate exists for the threshold value of the signal attribute, the adaptive procedure can utilize a larger step size at the start of the experiment to allow the threshold region to be reached in a smaller number of trials. The accuracy of the procedure can then be increased

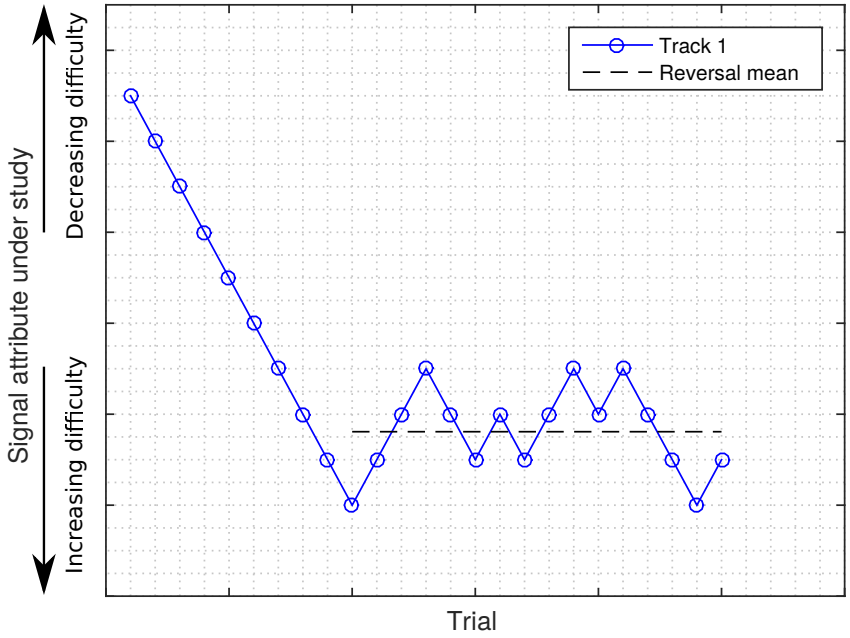


Figure 8: Example results from a generic up-down adaptive procedure with constant step size. The y-axis represents the magnitude of the signal attribute under study. The x-axis represents trial number. The signal attribute is adjusted according to the test subject's performance to steer the signal attribute towards a threshold value.

by decreasing the step size for trials in the convergence region, thus allowing the threshold to be estimated with greater resolution. The test procedure can be made more robust by requiring a larger number of consequent correct responses before adjusting the signal attribute. This increases the length of the experiment, but reduces the possibility of registering false positives.

3.2 Noise stimuli

As psychoacoustics is essentially the study of the input-output relationship between physical acoustical stimuli and the cognitive response of the auditory system, the properties of the input stimulus need to be well defined for a meaningful examination of the resulting output. While there are a wide variety of standard test signals used in hearing research (see e.g., [8] or [36] for an overview), for the purposes of this work the introduction of two commonly employed noise stimuli will suffice; namely white and pink noise.

White noise is characterised by a constant power spectral density across the audio band, resulting in a flat power spectrum as shown in the figure on the left hand side of Fig. 9. The power spectral density of pink noise on the other hand is inversely proportional to frequency, resulting in a power spectrum with -3 dB per octave decay characteristic as shown in the figure on the right hand side of Fig. 9. Whereas the $1/f$ spectral decay characteristic of pink noise ensures constant average signal power per

octave, in the case of white noise stimuli, the signal power per octave increases with frequency due to the constant power per hertz characteristic of the power spectrum. [37]

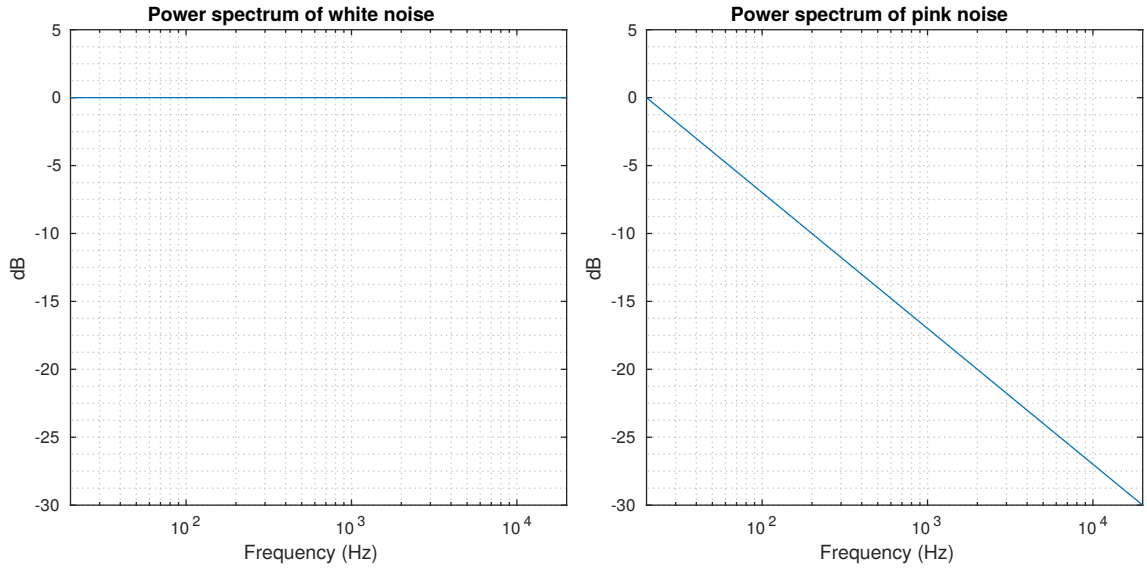


Figure 9: Power spectra of white and pink noise. White noise is characterised by a flat power spectrum yielding constant power per Hz. The power spectrum of pink noise has a $1/f$ decay characteristic.

3.3 Analysis of variance

The analysis of variance (ANOVA) is a standard statistical analysis procedure used to assess whether statistically significant differences exist between data sets. In psychoacoustic applications, this usually translates to testing the significance of some parameter of an acoustic stimulus on auditory perception. The core idea of ANOVA is to compare the magnitude of the variances present within data sets gathered under different treatment conditions, to the variances present between the means of these data sets. If the variance in treatment means is not sufficiently high compared to the variance present within the treatment groups, variations in the group means are likely to result from the within group variances rather than treatment effects. The following pages describe the process of computing a standard one variable ANOVA as outlined by Bech and Zacharov in [1].

The null-hypothesis H_0 of the basic one variable ANOVA procedure is that the means μ_k of the k groups under test are equal. The alternative hypothesis H_a contradicts the null hypothesis, i.e at least one of the means differs from the others within the tested population.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

$$H_a : \mu_i \neq \mu_j, \text{ for at least one combination of } i \text{ and } j.$$

The condition for rejecting the null hypothesis is that the variance of the group means between the k groups of the tested population is high enough that it cannot be explained by the variances of the n observations within the groups.

A one variable analysis of variance can be carried out on a population consisting of k independent groups, each containing some number nk of observations as depicted in Table 1.

Observation	Group 1	Group 2	Group 3	...	Group k
1	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	$\cdots$	$y_{k,1}$
2	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	$\cdots$	$y_{k,2}$
3	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	$\cdots$	$y_{k,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	$y_{1,n1}$	$y_{2,n2}$	$y_{3,n3}$	$\cdots$	$y_{k,nk}$

Table 1: Data matrix for one-way ANOVA.

A sum-of-squares figure SS_w can be used to characterise the total deviation from the group mean μ_k within each of the k groups. The figure is obtained by summing the squared deviations from the group mean.

$$\mathrm{SS}_w = \sum_{i=1}^{nk} (y_{k,i} - \mu_k)^2 \tag{2}$$

A similar figure SS_b can be calculated for the variance between groups by computing the sum of the squared deviations of the group means μ_k from the global mean $\overline{\mu}$ weighted by the number of observations nk within that group.

$$\mathrm{SS}_b = \sum_{i=1}^{k} nk(\mu_i - \overline{\mu})^2 \tag{3}$$

These sum-of-squares figures are then used to estimate the variances within and between groups. The variances are estimated by mean squares figures MS_w and MS_b obtained by dividing the sum of squares figures SS_w and SS_b by the within- and between groups degrees of freedom, df_w and df_b:

$$\mathrm{MS}_w = \frac{\mathrm{SS}_w}{df_w} = \frac{\mathrm{SS}_w}{N - k} \tag{4}$$

$$\mathrm{MS}_b = \frac{\mathrm{SS}_b}{df_b} = \frac{\mathrm{SS}_b}{k - 1} \tag{5}$$

where N denotes the total number of observations contained in the tested population [1]. The ratio of the mean squares figures define an F-ratio F

$$F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w}. \tag{6}$$

The obtained F-ratio is then checked against a cut-off score F_α of the F-distribution corresponding to the degrees of freedom used in computing the F-ratio and the

applied significance level; it is common practice to select a significance level of 5%. The required condition for rejecting the null hypothesis is then

$$F \geq F_\alpha(df_b, df_w). \tag{7}$$

The relationship between the F-ratio and its p-value, F_α and the test level is illustrated in the example distribution shown in Fig. 10 below. An example of an ANOVA table summarizing the numerical results of the analysis procedure is shown in Table 2.

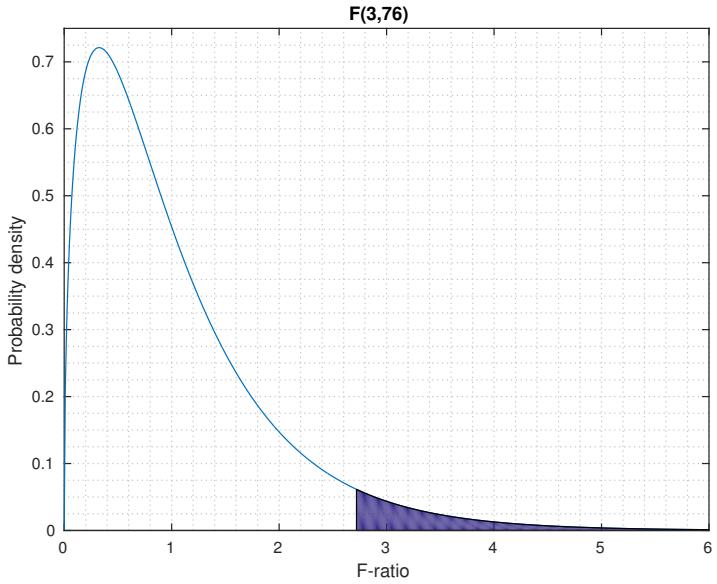


Figure 10: F-distribution F(3,76). The filled in area under the curve represents 5% of the total area under the curve. An F-ratio within this region fulfills the condition set by Eq.(7) and allows the null hypothesis to be rejected at 5% test level. F_α corresponds to the lowest F-ratio within the filled in region of the curve; in this case $F_\alpha = 2.73$. The significance value p corresponds to the percentage of the total area of the F-curve lying to the right of the tested F-ratio.

Dependent variable	Sum of squares	Degrees of freedom	Mean Square	F-ratio	Significance
Between-groups	SS_b	df_b	MS_b	F	p
Within-groups	SS_w	df_w	MS_w		
Total	$\mathrm{SS}_b + \mathrm{SS}_w$	$df_b + df_w$			

Table 2: A generic ANOVA table.

3.3.1 Auxiliary tests

The ANOVA-procedure relies on a set of assumptions about the test data that allow the involved computations to be as simple as shown above. The reliability of the analysis is compromised if these underlying assumptions are not met and a series of pre-tests should be performed to ensure that ANOVA can be conducted.

The basic assumptions of ANOVA are that the test data are normally distributed, have homogeneous variance and the gathered observations are independent from one another. While the last condition is mainly determined by experiment design and no reliable methods for testing the condition exist, standard statistical tests have been developed for testing the normality and variance assumptions.

Shapiro-Wilk test. The Shapiro-Wilk test is a commonly employed normality test recommended for sample sizes smaller than 50 [12]. The null- and alternative-hypotheses of the Shapiro-Wilk test are:

$$H_0 : \text{Data is normally distributed}$$

$$H_a : \text{Data is not normally distributed.}$$

Significance values larger than the chosen test level α therefore indicate no statistically significant deviations from normality.

Although ANOVA-procedures are fairly robust against small deviations from normality [14], it is deemed good practice to ensure that the data is not in excessive violation of the normality assumption. Slight skewness in the treatment group's distribution will not distort the ANOVA results significantly.

Levene's test. An often used method for testing the assumption of homogeneous variance is the Levene test. The null- and alternative-hypotheses are:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = ... = \sigma_k^2$$

$$H_a : \sigma_i^2 \neq \sigma_j^2, \text{ for at least one combination of } i \text{ and } j$$

Significance values below the test level therefore indicate unequal variances between the treatment groups. The test is essentially a standard ANOVA-test applied to the absolute deviations d_{ji} from the corresponding group medians $\tilde{y_j}$

$$d_{ij} = |y_{ji} - \tilde{y_j}| \begin{cases} i = 1, 2, 3..., n \\ j = 1, 2, 3..., k \end{cases} \tag{8}$$

The equality of the means of the deviation groups is then assessed according to the ANOVA procedure outlined above. If no statistically significant differences exist between the means of the deviation groups, the same is true for their variances and the homogeneity assumption is therefore valid [7]. If the data fails Levene's test for equal variances, a Welch's correction should be applied to the F-statistic for robust testing [14]. An analysis of variance using a Welch-corrected F-statistic is referred to as a Welch-ANOVA. Welch-ANOVA is robust against deviations from the assumption of homogeneous variances for data sets containing both equal and unequal numbers of observations within the treatment groups [22].

Tukey's test. Given that a successful rejection of the ANOVA null hypothesis merely implies that at least one pair of group means are unequal to a statistically significant degree, a series of post hoc tests needs to be carried out for closer analysis. A standard post hoc test for the pairwise comparison of group means is Tukey's test. Its null- and alternative hypotheses are

$$H_0 : \mu_i = \mu_j$$

$$H_a : \mu_i \neq \mu_j$$

where i and j denote the indices of the two treatment groups under test. The threshold of statistical significance T_α is derived from the Studentized range statistic $q_\alpha(df_b, df_w)$ corresponding to the degrees of freedom df_b & df_w associated with the data set. If the number of observations is equal in all pairs of treatment groups the threshold is of the form

$$T_\alpha = q_\alpha(df_b, df_w)\sqrt{\frac{\mathrm{MS}_w}{n}}, \tag{9}$$

where MS_w is the within groups variance as computed in the analysis of variance and n is the number of observations in the groups [7]. If the groups contain an unequal number of observations, the threshold of significance $T_{\alpha i,j}$ for groups i and j is of the form

$$T_{\alpha i,j} = \frac{q_\alpha(df_b, df_w)}{\sqrt{2}}\sqrt{\mathrm{MS}_w(\frac{1}{n_i} + \frac{1}{n_j})}, \tag{10}$$

where n_i and n_j correspond to the number of observations in the two groups currently under comparison [7]. If the groups are of unequal size, the procedure is referred to as the Tukey-Kramer test.

The test compares all $i \neq j$ pairs of groups and rejects the null-hypothesis if the absolute value of the difference between the means of the groups under comparison fulfill the condition

$$|\mu_i - \mu_j| \geq T_{\alpha i,j}. \tag{11}$$

In Chapter 6, these tools are applied to the results obtained from the conducted listening test.

3.4 Masking

Masking refers to the common phenomenon where the presence of one acoustic signal disturbs the perception of another by raising the threshold of audibility of the masked sound. The effect can be quantified by determining the magnitude of the threshold-increase in dBs. Masking is most efficient at frequencies where the spectra of the two signals overlap. In such a situation, both signals will excite the same regions on the basilar membrane and thus stimulate the same sets of hair cells in the cochlea. The vibrations caused by the stronger signal will tend to dominate the output of the hair cells and the information carried by the weaker signal is lost. Masking becomes more efficient as the spectral overlap between the two signals increases; consequently, pink and white noise are typically used to provide

efficient wide band masking in laboratory conditions. Spectral masking can be better understood through the concept of auditory filters introduced in Section 3.5.

The masking effect is not limited to situations where the two signals are presented simultaneously. Rather, the effect is active for short periods of time both before and after the presentation of the masking signal. Masking that occurs outside the temporal window during which the masker signal is present is referred to as temporal masking. It can be divided into pre- and post-masking as shown in Fig. 11. Pre-masking occurs during the time period preceding the onset of the masker; its effective time window is limited to 5-10 ms before the onset of the masker [17]. Post-masking is a similar—albeit much stronger—phenomenon that occurs after the offset of the masker. This effect can be active for up to 200 ms following the termination of the masker signal. The amount of masking depends on the frequency content, amplitude and duration of the masker. Post-masking is related to ringing effects on the basilar membrane—i.e., the membrane continues to vibrate for a short period of time even after the offset of the stimulus— and the non-instantaneous adaptation of neural mechanisms [30].

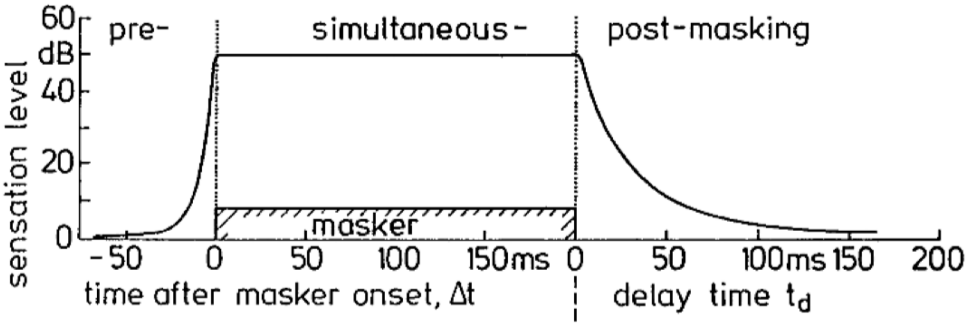


Figure 11: Illustration of the three temporal regions of the masking phenomenon. Pre-masking refers to the masking effect occurring in the short period of time before the onset of the masker signal. Simultaneous masking refers to the situation where both of the signals are active at the same time. Post-masking occurs after the offset of the masker signal. Adopted from [8].

3.5 Auditory filters and the equivalent rectangular bandwidth

The auditory system is able to separate frequency components in a wideband stimulus due to the mechanical frequency separation performed by the basilar membrane. The minimum required frequency difference for two tones to be distinguishable from one another is set by the critical bands centered at the frequencies in question. The critical band is a central concept in auditory research and refers to a band of frequencies the auditory system processes as one unit when frequencies within the band are presented to the cochlea simultaneously. Critical bands can be conceptualized as the pass bands of band-pass filters known as auditory filters. The measurement of their shape and width has been a topic of continuous research since Fletcher first introduced the concept of the critical bandwidth in the 1940s [10]. For the purposes of this work, a measurement method of special interest is the notched-noise method introduced

by Patterson in [32]. The measurement essentially consists of a set of threshold measurements for a tone located in the middle of a band-reject filtered white noise masker as shown in Fig. 12. The width of the notch is varied and a corresponding threshold is determined for each notch width. This set of threshold values is then used to deduce the width and steepness of the skirts of the filter centered at the test tone frequency through a polynomial fitting procedure. The resulting auditory filters obtained with the standard notched-noise method are symmetric about the center frequency as shown in Fig. 12.

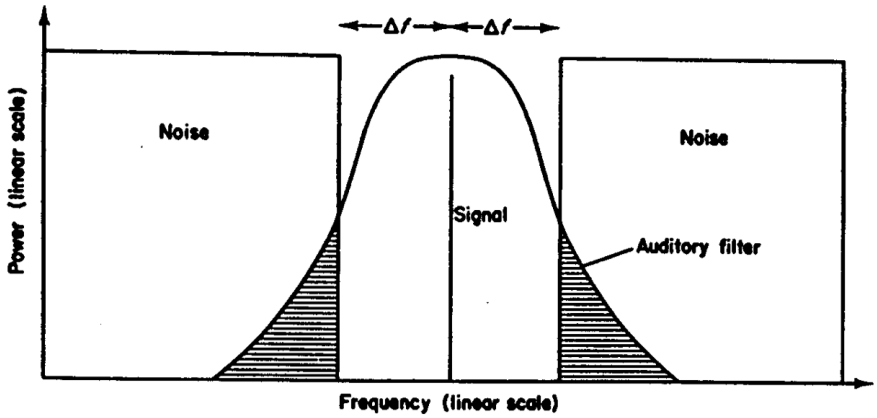


Figure 12: The notched-noise measurement arrangement. Adopted from [30].

The width of an auditory filter is a function of its center frequency. Auditory filters at high center frequencies are wider and less sharply tuned than those at low center frequencies implying poorer frequency resolution at the high end of the audio band [30]. Because the shape of the filters tends to vary with frequency, a convenient means of comparing the bandwidths of auditory filters regardless of their shape is to normalize the shape of the filters to a common reference. An often used method is to use the equivalent rectangular bandwidth representation (ERB) of the filter, wherein the filter is converted to an ideal band-pass filter with infinitely steep skirts and a flat pass-band as shown in Fig. 13 below. The ERB-conversion of parametrized auditory filters derived with notched-noise measurements is described in detail in [13]. For the purposes of this chapter however, it will suffice to say that the peak value of the ERB-filter is fixed to that of the measured filter and the rectangular filter's width is adjusted so that both filters pass the same total power when a flat spectrum noise signal is applied to the input [37].

3.5.1 ERB-rate scale

Due to the fact that the relationship between physical stimuli and sensory responses is generally non-linear, it is often convenient to map physical units to a scale that reflects the properties of the sensory system. In the case of auditory perception, a common procedure is to map frequency expressed in hertz to an auditory frequency scale that relates physical frequency to the frequency resolution ability of the cochlea; the ERB-rate scale is one such scale. It defines an auditory scale based on the

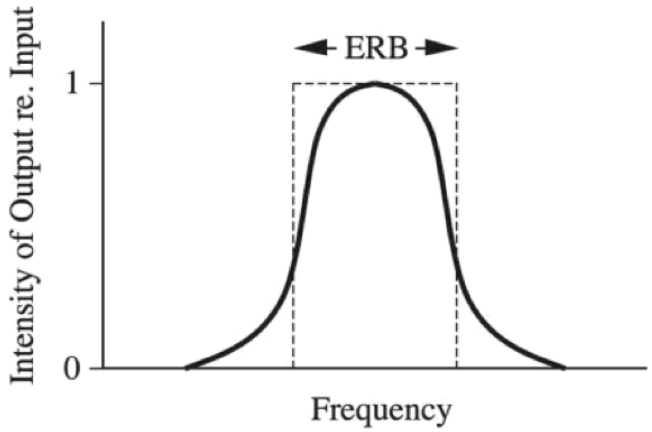


Figure 13: An generic auditory filter and its ERB-presentation. Both filters pass the same amount of signal energy when excited with a white noise stimulus. Adopted from [35].

equivalent rectangular bandwidth of auditory filters measured with the notched-noise method. The scale is constructed so that one unit (z_{ERB}) on the auditory scale corresponds to a frequency increment equal to the ERB of one auditory filter [13]. A z_{ERB} number therefore denotes the number of auditory filters below the frequency corresponding to that z_{ERB} number. The relationship between physical frequency expressed in hertz and the ERB-rate scale is expressed by the equation [36]

$$z_{ERB} = 21.3\log_{10}(1 + f/228.7). \tag{12}$$

This non-linear mapping is visualized in Fig. 14 below. Note that one unit on the auditory scale corresponds to a wider bandwidth in Hz as the z_{ERB} number increases; this reflects the widening of auditory filters at high frequencies.

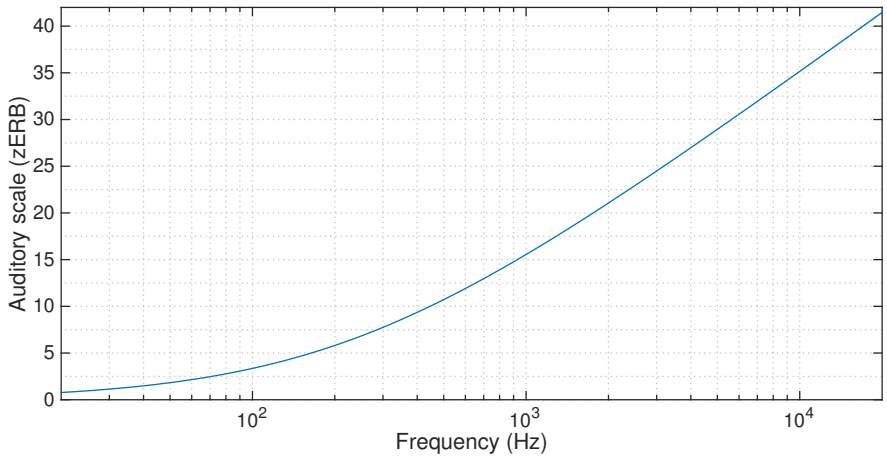


Figure 14: ERB-rate scale as a function of audio band frequencies, as calculated by Eq.(12).

4 Localization

Auditory localization refers to the process through which the hearing system forms the perception of the apparent spatial position of sound sources in three-dimensional space. Sound sources are localized through cues provided by multiple auditory mechanisms that contribute to the final perceived source location. This chapter begins by introducing some basic terminology and the standard coordinate system employed in most localization studies. The rest of the chapter is devoted to introducing the basic binaural and monaural localization mechanisms involved in static and dynamic localization of sound sources under anechoic conditions.

4.1 Double pole coordinate system

Localization studies in three-dimensional space often employ the spherical coordinate system shown in Fig. 15 that allows source directions to be uniquely defined by a single pair of angles; this arrangement is generally referred to as the double pole coordinate system. The horizontal direction component is defined by the azimuth angle θ rotating about the z-axis of the conventional Cartesian coordinate space. Similarly, the vertical component is defined by the elevation angle ϕ rotating about the y-axis of the Cartesian coordinate space.

In spatial hearing applications, the coordinate space is fixed with respect to the observer's head, by placing the origin of the coordinate system at the midpoint of the observer's interaural axis. The planes defined by the conditions y = 0 and z = 0, intersecting at the origin of the coordinate system are referred to as the median and azimuthal planes. The plane defined by the condition x = 0 divides the coordinate space into frontal and rear hemispheres. Similarly, the median plane splits the coordinate space into left and right hemispheres and defines a plane equidistant to both of the observer's ears. Finally, the azimuthal plane divides the three-dimensional space into upper and lower hemispheres.

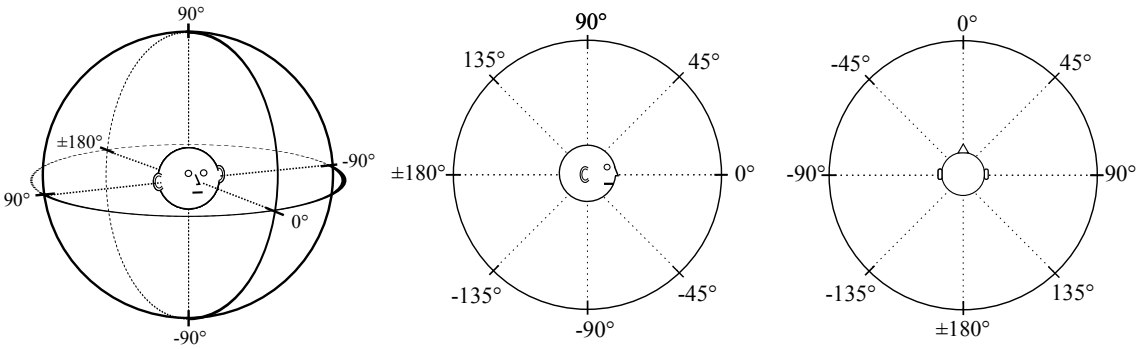


Figure 15: An illustration of the double pole coordinate system used in spatial sound applications.

4.2 Interaural differences

The classical binaural theory of sound localization was originally proposed by Lord Rayleigh at the beginning of the twentieth century [39]. Rayleigh's model explains human sound localization through two processes relying on interaural differences in signal parameters and was hence dubbed the duplex theory of localization. The localization mechanisms can be understood through a geometrical examination of head related acoustics.

4.2.1 Interaural time difference

In a simplistic but convenient model, the human head can be represented as a sphere with the two ears located on opposite poles of its equatorial plane. The difference in arrival times—or the interaural time difference (ITD)—at two such poles for a sound wave propagating at some velocity c, emitted by a source at an angle θ in azimuth can be expressed as [29]

$$\text{ITD} = \pm(r/c)(\theta + \sin(\theta)) \tag{13}$$

where r is the radius of the sphere. Typically the interaural path difference produces ITD values ranging from zero to approximately 650 µs for an average sized human head. An example of theoretical ITDs for azimuth angles -90° to 90° as predicted by equation (13) are shown below in Fig. 16. Maximum ITD values occur at azimuth angles of ±90° and the minima are located at 0° and ±180°.

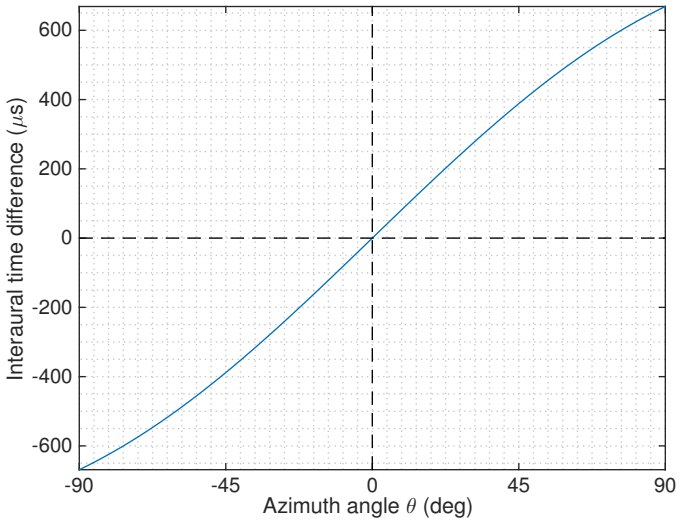


Figure 16: Interaural time difference as a function of azimuth angle for a spherical head model of radius 9cm.

The auditory system's sensitivity to changes in ITD is a function of several signal parameters including ITD, level, temporal envelope and the frequency content of the stimulus. Variations in ITD are more easily detected when the original ITD value

corresponds to source locations near the median plane. Similarly, sensitivity to ITD is greater for broadband stimuli. The term *just noticeable difference* (JND) is a term often used to refer to the smallest detectable change in a signal parameter. In the case of ITD, the JND under optimal stimulus conditions has been reported to vary between 10 to 20 µs. [36]

4.2.2 Interaural level difference

The interaural level difference is defined as the sound pressure level difference between the two ears. Due to the physical phenomenon of diffraction, the human head does not present an acoustical obstacle to sound waves with a wavelength comparable to the dimensions of the head. At higher audio frequencies the wavelength of sound is significantly smaller and the dimensions of the head are large enough in comparison for the head to act as a reflective obstacle to incident sound waves. As a result, an interaural sound pressure level difference can be formed between the two ears if the head casts an acoustic shadow over either of the two ears. This effect is illustrated in in Fig. 17. The hearing system takes advantage of this phenomenon in the localization of sound sources with high-frequency components present in the source spectrum. Because the shadowing effect of the head becomes more prominent at high frequencies, the interaural level difference is a function of both frequency and source direction as illustrated by Fig. 18.

Figure 17: Illustration of the frequency dependence of interaural level difference. A plane wave propagates from left to right. In the figure on the left hand side, the wavelength λ is larger that the dimensions of the head and no acoustic shadowing occurs when the wave encounters the observer's head. In the figure on the right hand side, the dimensions of the head are large relative to the wavelength λ and the head presents an acoustical obstacle that leads to a sound pressure difference between the two ears.

As in the case of interaural time differences, the auditory system's sensitivity to changes in ILD is a function of several stimulus parameters including ILD and frequency. The JND of ILD is smaller for stimuli with original ILDs corresponding to source locations near the median plane. Similarly, changes in ILD are detected more easily at low frequencies. The JND of ILD has been reported to be as low as 2 dB under optimal stimulus conditions. [36]

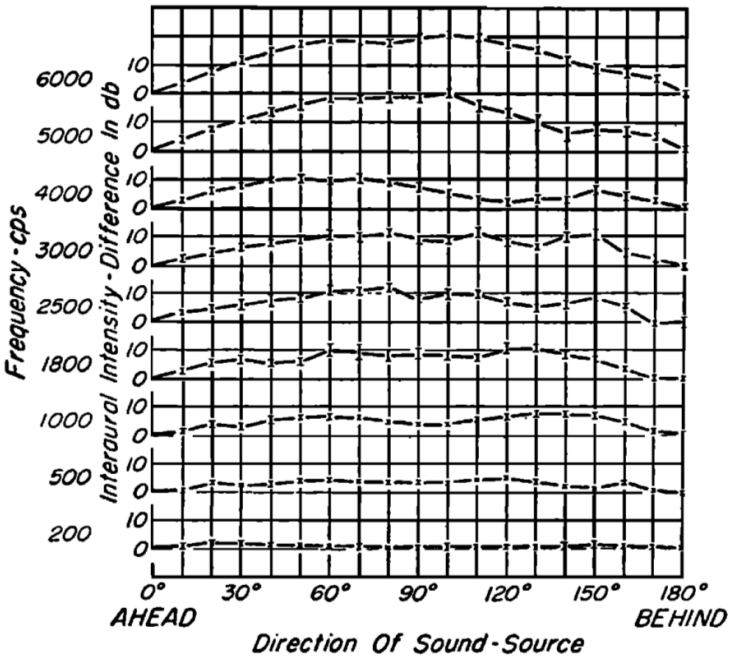


Figure 18: Interaural level difference at different frequencies for sources at various azimuth angles. Sound pressure difference is at its highest for source locations coinciding with the interaural axis and increases with frequency. Adopted from [9].

4.2.3 Ambiguity of interaural cues

The binaural cues are fundamentally limited by the geometrical ambiguity of the interaural path difference. Using the simplified two-dimensional model shown on the left hand side of Fig. 19 to represent the geometry of sound waves arriving at the two ears, it can be shown that a given path difference specifies a parabolic locus of possible source locations rather than a unique coordinate pair. Furthermore, extending the model to three dimensions increases the span of possible source locations to a conical surface defined by rotating the parabola about the interaural axis as shown in the figure on the right hand side of Fig. 19. The conical locus defines the set of possible source locations corresponding to a given interaural difference value; this locus is referred to as the cone of confusion [3]. The localization accuracy of the human hearing system however is much more precise than predicted by the cone of confusion; localization accuracy is improved beyond the limits set by the cone of confusion through processes employing additional cues described in the following sections.

4.3 Spectral cues

A major weakness of the classical duplex theory is its inability to account for the hearing system's ability to discriminate between sound sources in the front and back sectors of the double pole coordinate system as the ITD and ILD cues are symmetrical about the interaural axis as defined by the cone of confusion. Similarly, sources in the median plane produce the same ILD and ITD cues regardless of elevation

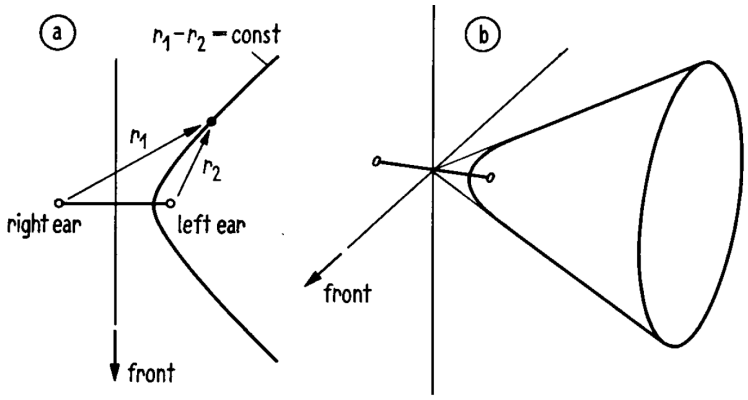


Figure 19: Illustration of the cone of confusion. The figure on the left hand side shows the parabolic locus of possible source locations implied by an arbitrary interaural difference value in a simplified two-dimensional model. The figure on the right hand side shows the model extended to three dimensions and the resulting conical surface of possible source locations. The interaural cues do not specify a unique source location but rather a conical locus of possible locations referred to as the cone of confusion. Adopted from [2]

angle. Despite these limitations, the hearing system is still able to discriminate between different source elevations; much of the information used in making elevation judgements is derived from the spectral pattern of the sound pressure signal present at the eardrum.

Spectral cues are derived from the reflection pattern formed in the cavities and ridges of the pinna. As these cues are formed by the interference patterns of direct and reflected sound waves, the effectiveness of spectral cues is limited to the high-frequency region, where the dimensions of the pinna are large enough in comparison to the wavelength of incident sound waves to present an acoustically reflective obstacle to the said waves. Likewise, for the auditory system to be able to identify and decode spectral features from the sound pressure signal presented to the ears, the stimulus needs to have a sufficiently wide bandwidth. Due to the small dimensions of the pinna cavities, the pinna cues require signal energy above approximately 4 kHz for these cues to become meaningful [35]; reflections from the shoulders and rest of the torso provide a longer reflection path that forms additional spectral features at frequencies below 2 kHz [20]. In addition to resonance and antiresonance cues formed in the cavities of the pinna, the external ear provides additional acoustic shadowing at high frequencies for sources located in the rear hemisphere. The head, pinna and shoulders therefore form a direction-dependent filter that modifies the sound pressure spectrum formed at the eardrum as a function of source direction. This spectral modification can be conceptualized as an acoustic transfer function between a given source location and the eardrum. These acoustic input-output relationships are known as head-related transfer functions (HRTF) and are directly measurable via an impulse response between a point source at a given direction and a test subject's eardrum in anechoic conditions. The obtained head-related impulse response can then be transformed to the frequency domain to yield the corresponding HRTF. Fig. 20 illustrates how each source direction yields a unique pair of HRTFs corresponding to the spectral modifications characteristic of that source direction.

The black and red plots correspond to the HRTFs measured at the right and left ears respectively. For sources coinciding with the median plane, the spectra measured at the two ears is nearly identical. Conversely, sources located off the median plane show discrepancies between the spectral pattern measured at the two ears. Furthermore the location of the spectral peaks and notches varies as a function of source location.

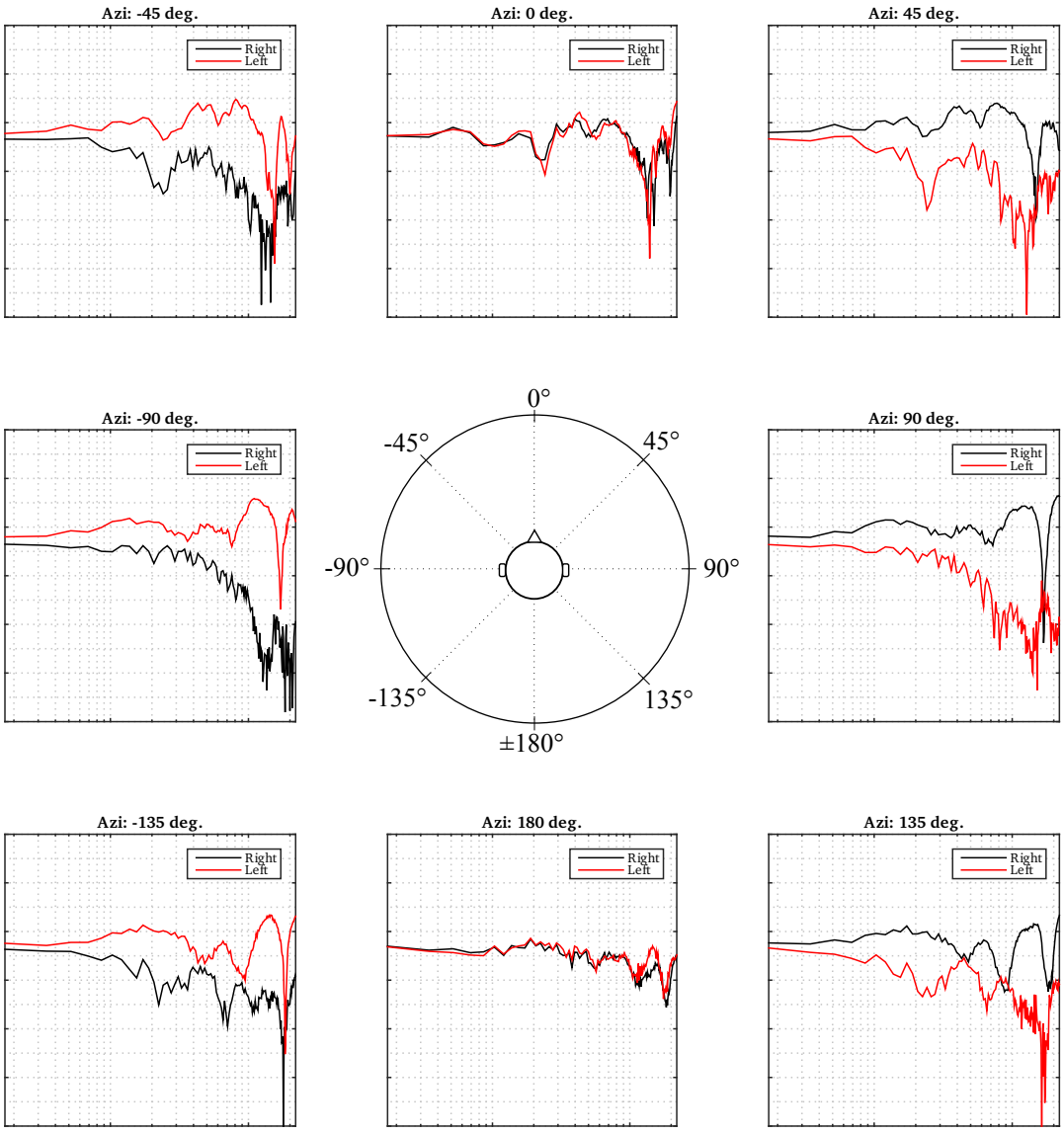


Figure 20: Illustration of head-related transfer functions for both ears measured at different angles on the azimuthal plane; each source direction defines a unique pair of acoustical transfer functions corresponding to the spectral pattern formed at the two ears. The x-axis of each plot represents frequency; the y-axis corresponds to magnitude in expressed in dB. Similar measurements can be made for elevated sources.

HRTFs provide a means of implementing virtual acoustics for headphone listening. By realizing HRTFs as digital filters, source directions can be simulated in headphone listening by passing signals through the binaural filters corresponding to the source

direction to be simulated. Such a procedure will replicate the sound pressure spectrum produced by a real source at the same direction and thus provide the auditory system with the spectral cues corresponding with a source in that direction.

4.4 Dynamic cues

The treatment of the previously described localization processes has been limited by the simplistic assumption of static sources and observers. However, in a real localization scenario this assumption is rarely justified, as more often than not, a human observer performs a series of head movements to facilitate the localization process. The additional information to source location furnished by head movements are referred to as dynamic cues.

Natural head movements can be characterised as combinations of rotations about three perpendicular axes centered at the middle of the head as shown in Fig. 21. Terms pivoting, tipping and rotating refer to rotations about the x-,y- and z-axes respectively. Naturally, these movements bring about a change in the binaural and spectral cues. Consider for example the situation shown on the left-hand side of Fig. 22 where a sound source is located at the intersection of the observer's azimuthal and frontal median planes. As binaural cues are symmetric about the interaural axis, front-back confusion is likely unless the source spectrum provides access to accurate spectral cues. Should it be that the source spectrum lacks these cues, the front-back ambiguity can still be resolved by head movements as the direction of the resulting change in the binaural cues depends on the location of the sound source [44],[46],[47]. Referring to the figure on the right-hand side of Fig. 22, it can be seen that as the observer rotates his head, the sound pressure level decreases at the left ear as the acoustic shadowing effect due to the head increases. Conversely, the sound pressure level increases at the right ear as it is brought closer to the sound source. Similarly, the signal arrives earlier at the right ear and later at the left ear producing a change in the ITD. If the source were in the rear hemisphere, the relationship between the changes in the binaural cues would be reversed. Pivoting movements produce similar changes in the binaural cues if the sound source is located above or below the azimuthal plane. Tipping results in no changes in the binaural cues as the interaural axis remains stationary during these movements. Tipping movements however change the sound pressure spectra formed at the eardrums due to the direction-dependent filtering effect of the pinnae and the torso. Experimental evidence suggests that low-frequency ITD is the salient interaural cue used in forming the dynamic cue; see e.g., [23], [24], [31], [33], [34], [40].

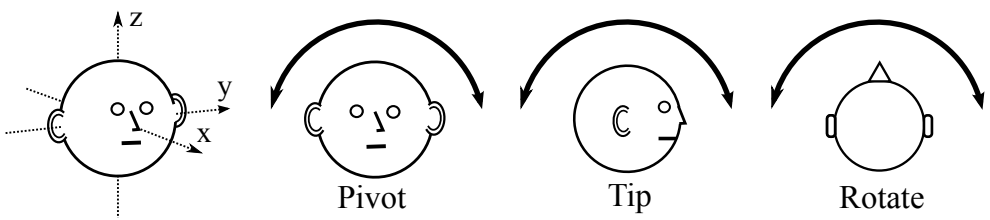


Figure 21: Schematic illustration of head movements about three axes. Figure adapted from [40]

The classical theory of dynamic localization cues was developed by Wallach in two classic papers [41], [42]. He proposed that dynamic cues are formed by integrating the changes in the apparent source location with the changes in head orientation information that take place during head movements. The classical treatment of dynamic cues is largely formed around a geometrical analysis of the relative changes occurring in the lateral angle γ and the angular displacement of the interaural axis during rotating head movements. Here the lateral angle refers to the angle measured between the interaural axis and the apparent source location as indicated by the binaural cues, i.e. the angle between the interaural axis and the surface of the cone of confusion [41]. While static binaural cues result in a number of possible source locations as defined by the cone of confusion, only a limited number of these locations result in an identical sequence of lateral angles when the observer moves his head. Wallach's theory therefore states, that the auditory system resolves the ambiguities of the cone of confusion by identifying the sequence of lateral angles presented during a given head rotation. The hemiplane ambiguity is resolved by the direction of change in the lateral angle as was illustrated in the example of Fig. 22; the remaining elevational ambiguity is resolved by monitoring the change in lateral angle produced by the rotation [41].

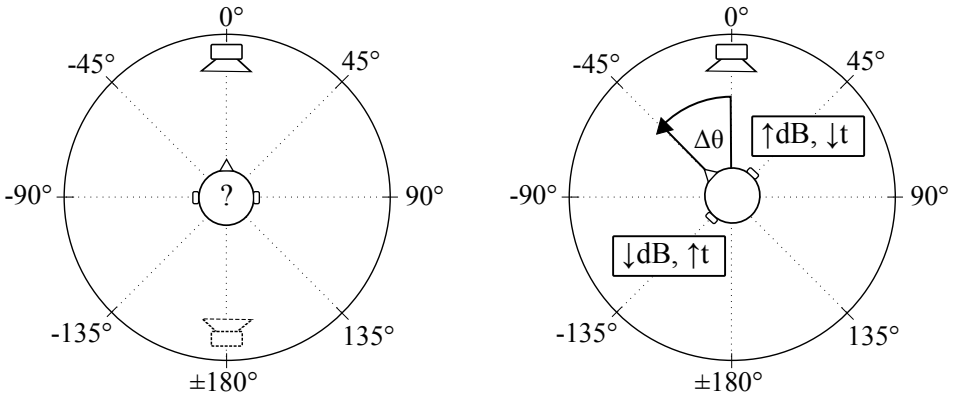


Figure 22: Example of how dynamic cues are used in solving front-back ambiguity. The initial situation is shown in the figure on the left-hand side. The binaural cues produced by the source at zero azimuth are identical to those produced by a phantom source in the rear hemiplane at ±180° azimuth. Dynamic cues resolve this ambiguity as shown in the figure on the right.

Geometrically, the lateral angle can be expressed as a function of azimuth and elevation angles by solving the appropriate direction cosine formula for the angle between the interaural axis and the unit vector pointing towards the source location [6]

$$\gamma = \cos^{-1}(\cos\phi\sin\theta). \tag{14}$$

Fig. 23 shows how the lateral angle changes as a function of azimuth angle for sources at various elevations. Note that the angle can be measured from either side of the interaural axis. Here the angle was measured from the right side of the interaural axis for convenience.

For a simplified illustration of the relationship between the change in lateral angle and head rotation, consider the three pairs of figures shown in Fig. 24 depicting

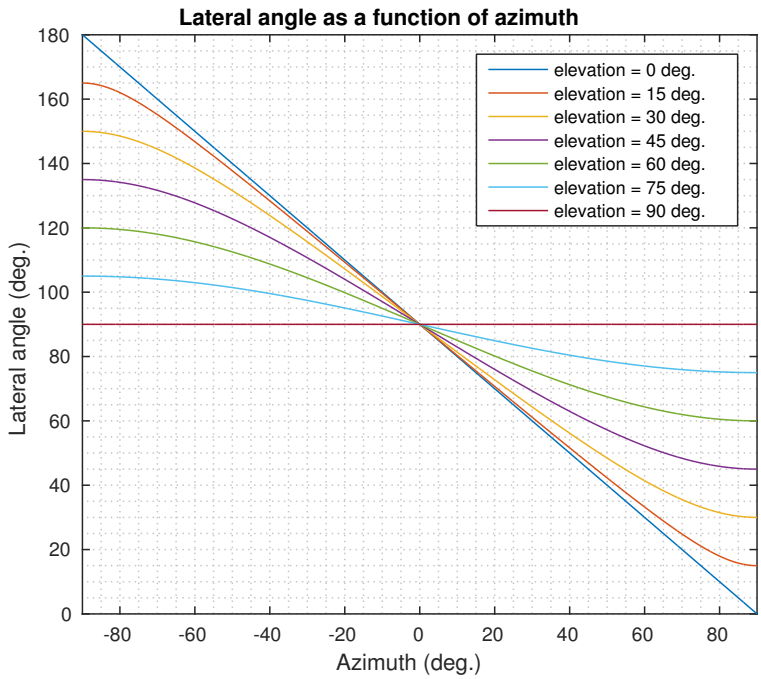
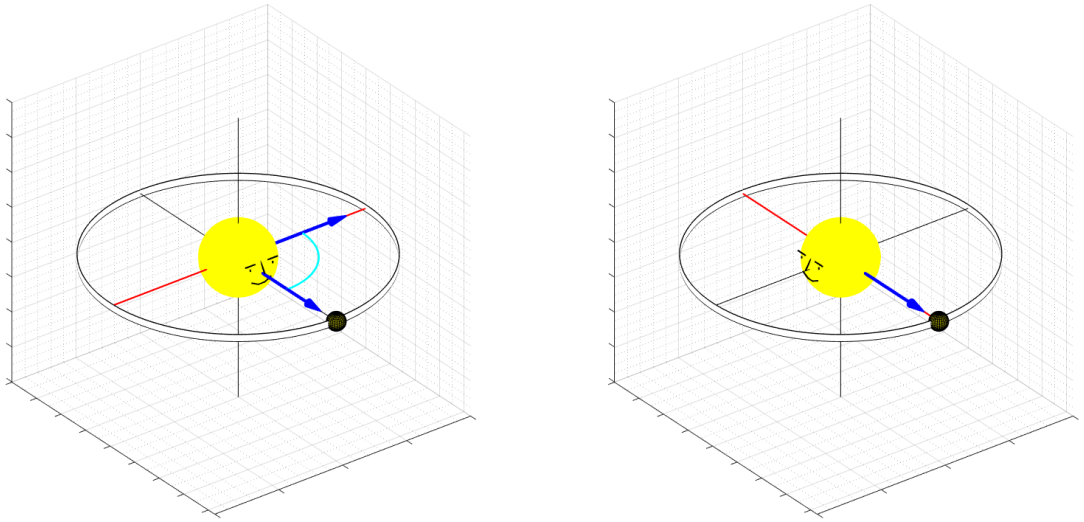
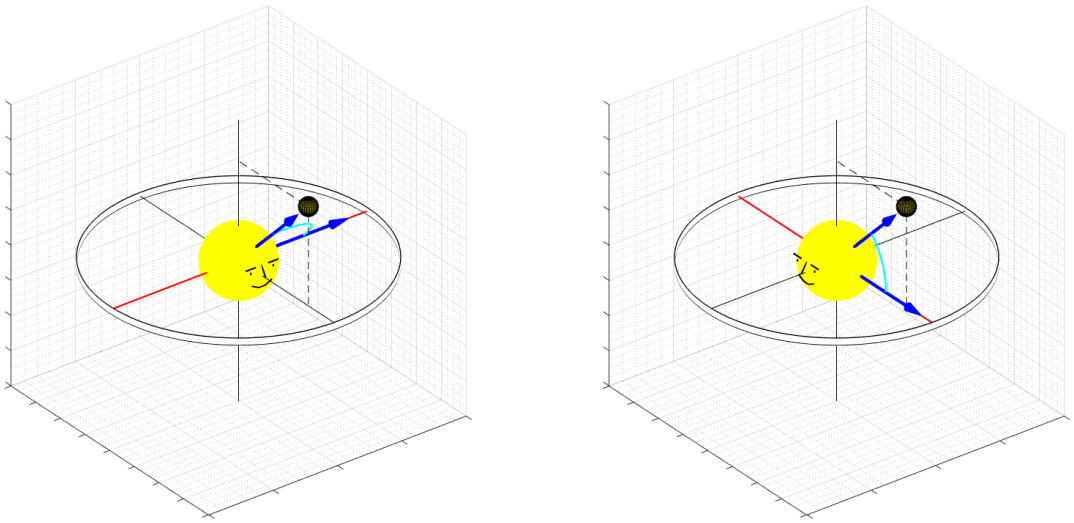


Figure 23: Lateral angle measured from the right hand side of the interaural axis as a function of azimuth at various elevations as predicted by Eq.(14). The rate of change in lateral angle is approximately linear near the median plane. The function is symmetric about zero azimuth.

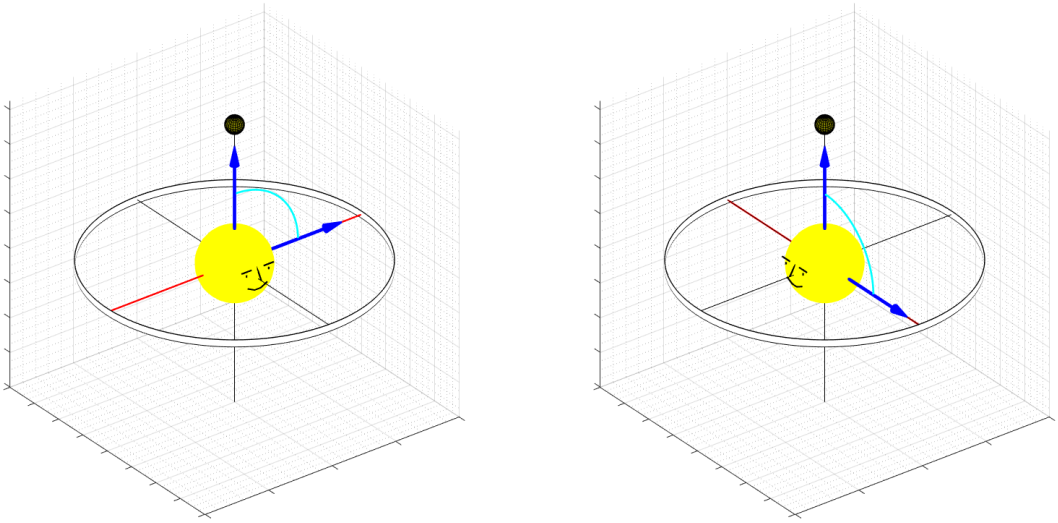
sound sources at elevations of 0, 45 and 90 degrees. The left and right hand sides of the three figures represent the spatial arrangement before and after the observer executes a 90-degree head rotation. The red line represents the observer's interaural axis. The angle between the blue arrows corresponds to the lateral angle. In the pair of figures at the top of Fig. 24, the head rotation causes the lateral angle to be reduced from 90 degrees to 0 degrees. The ratio of the change in lateral angle and angular displacement of the interaural axis is therefore equal to $90/90 = 1$. This relation holds for all source azimuths at zero elevation. In the pair of figures at the middle of Fig. 24, the source is elevated at 45 degrees in the median plane of the observer. The lateral angle in the initial situation is again 90 degrees as measured on the oblique plane defined by the blue arrows. A 90-degree rotation moves the interaural axis below the sound source, resulting in a lateral angle of 45 degrees. In this case, the ratio of the change in lateral angle and angular displacement of the interaural axis is therefore equal to $45/90 = 0.5$. Finally, refer to the figures at the bottom of Fig. 24 where a sound source is elevated above the observer at 90 degrees. A 90-degree rotation results in no changes in the lateral angle. The ratio of lateral angle variations to rotation angle in azimuth is therefore $0/90 = 0$ for sources elevated at ± 90 degrees. Note that the rate of change does not indicate whether the source lies on the observer's upper or lower hemisphere. Natural head movements however are rarely confined to rotations about a single axis and the small naturally occurring pivoting movements associated with head rotations resolve the hemisphere ambiguity even in the absence of spectral cues.



(a)



(b)



(c)

Figure 24: Illustration of changes in lateral angle resulting from 90 degree rotations for three different source elevations. The red line denotes the interaural axis and the black sphere represents the sound source. The lateral angle is measured between the blue arrows.

4.4.1 Ambiguity of dynamic cues

Despite being able to resolve source locations on the cone of confusion, dynamic cues can themselves provide ambiguous information in situations where the sound source is not stationary during head movements. Consider for example the situation illustrated in Fig. 25 where a sound source located on the observer's azimuthal plane is moving in azimuth at twice the observer's rotation rate. Such a situation results in a sequence of lateral angles that has two possible interpretations: a moving source in the frontal hemiplane or a stationary source in the rear hemiplane. It has been proposed that the auditory system generally prefers the interpretation that minimizes source movement; this tendency is referred to as the principle of least displacement [42].

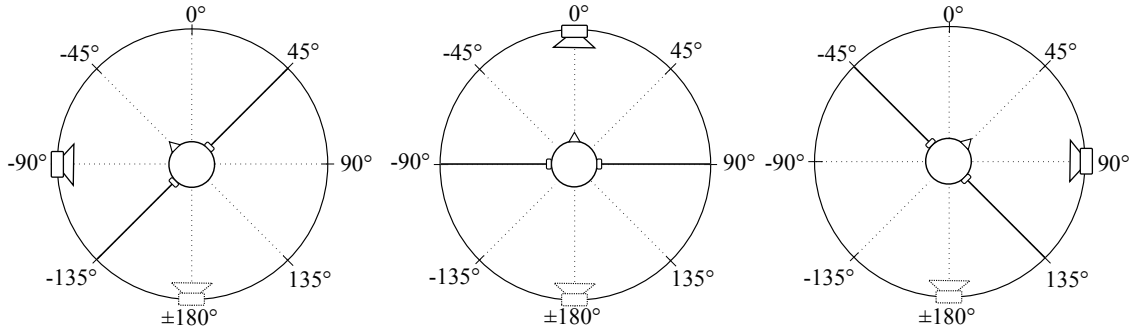


Figure 25: Ambiguity of dynamic cues when the sound source is in motion during a head rotation. Here the angular coordinates of the plots are fixed rather than referenced to the observer's head. The figure on the left shows the initial situation where the sound source is located at -90 degrees and a phantom source produced by the ambiguous primary cues is mirrored to ±180 degrees about the interaural axis. The figures in the middle and on the right show the situation as the observer rotates in azimuth and the source is displaced in the same direction at twice the observer's rotation rate. The phantom source is always at the same location. Image adapted from [23].

It should then be possible to produce synthetic source locations by artificially presenting a sequence of lateral angles to an observer during head movements. The principle of least displacement predicts, that if the presented sequence of lateral angles can be interpreted as a stationary source, such an interpretation will be preferred over the actually presented moving source interpretation. As a simple example of this, refer back to the bottom of Fig. 24, where an overhead source results in a static lateral angle of 90 degrees. According to the principle of least displacement, the same apparent source direction could be produced synthetically by moving a sound source so that it remains at zero degrees azimuth and elevation at all times during the observer's head rotation. Such a head-orientation-coupled panning procedure would result in a static lateral angle of 90 degrees, thus producing no changes in lateral angle for a given displacement of the interaural axis; a characteristic of sources elevated at ±90 degrees.

More generally, the synthetic elevation produced by head-motion-coupled azimuthal displacement of unelevated sound sources can be characterised by a displacement factor $\frac{\Delta\alpha}{\Delta\beta}$, where $\Delta\beta$ denotes the angular displacement of the interaural axis in azimuth and $\Delta\alpha$ is the corresponding angular displacement of the sound source. The

synthetic elevation v produced by a given displacement factor can be approximated by the equation [41]

$$v = \cos^{-1}(1 - \frac{\Delta\alpha}{\Delta\beta}). \tag{15}$$

This relation is plotted in Fig. 26 below. Referring to the plot, it can be seen that the equation predicts displacement factors of 1 and 2 resulting in synthetic source directions at 90 and 180 degrees elevation, corresponding to an overhead source and hemiplane reversal; this is consistent with the two examples presented previously in this section.

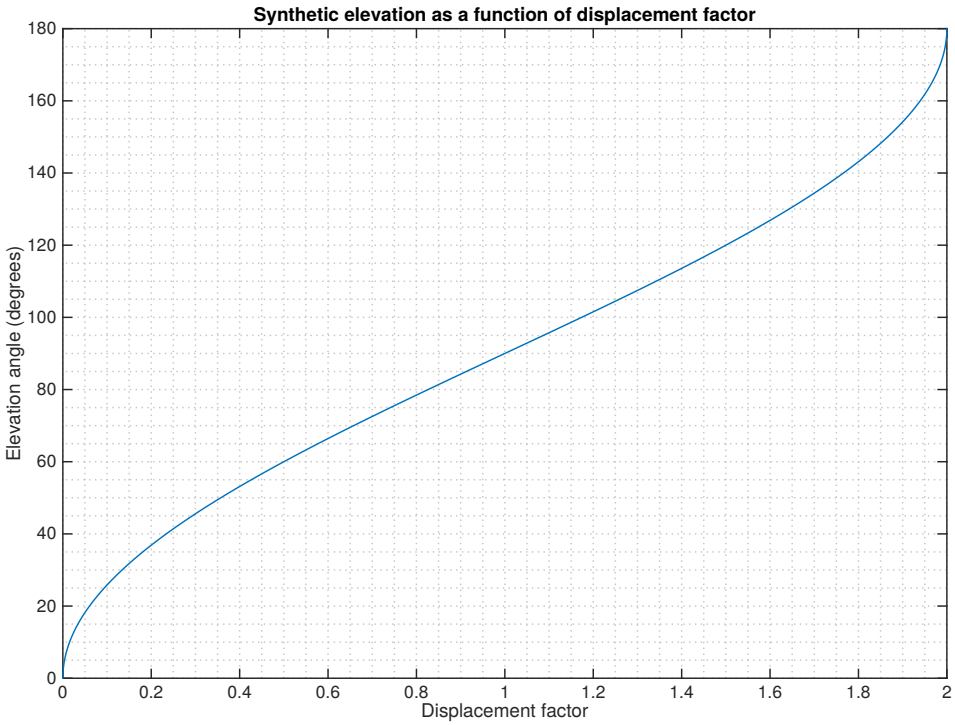


Figure 26: Synthetic elevations produced by displacement factors 0-2 as predicted by Eq. (15).

Because a synthetic sequence of lateral angles can be chosen so that the least displacement interpretation of dynamic cues implies a source location that differs from the location indicated by spectral cues, the principle of least displacement provides a useful means for weighing the relative salience of these two localization processes. By setting the two cues in conflict with one another so that the implied source locations are mutually exclusive, the active process used in localization can be easily identified.

4.4.2 Sensory information required in forming the dynamic cue

In forming the dynamic cue, the auditory system compares the rates of change in the lateral angle and head orientation. The auditory processes—namely: the binaural cues—involved in forming the lateral angle were presented in a Section 4.2. In this

section, a condensed overview of studies assessing the origin of the head orientation information is presented for the sake of completeness.

In [42], Wallach proposed three sets of sensory data that could provide the auditory system with the head orientation information required for forming the dynamic cue: proprioceptive information from the muscles of the neck that are active during head movements, vestibular information from the organ of balance and visual information provided by the eyes. He performed a series of experiments that tested the salience of dynamic cues under different test conditions that denied access to different combinations of the three proposed forms of sensory data. These experiments are briefly described below.

In the experimental scenario common to the two experiments described in the following paragraphs, subjects were seated on a rotating chair in the middle of a circular loudspeaker arc as shown in Fig. 27. A mechanical commutator switch was attached to the test subject's head that allowed a musical stimulus to be routed to different loudspeakers depending on the azimuthal orientation of the subject's head. Synthetic source locations could then be produced as estimated by the theory of dynamic cues by presenting the subject with an appropriate sequence of lateral angles during head rotations. The effectiveness of the dynamic cue under different test conditions could then be assessed based on whether the test subject perceived the synthetic source location or not.

In the first series of experiments, participants performed head rotations with access to all three of the possible sources for head orientation information; reportedly all five subjects perceived the source in the direction corresponding to the least displacement interpretation of dynamic cues. In the second series of experiments, the vestibular information was isolated by blindfolding the subject and having the experimenter rotate the chair the subject was seated in. The subject was therefore under passive rotation without visual or proprioceptive stimulation, thus leaving the stimulation of the vestibular organ as the only active source of information for head orientation. A synthetic source direction was produced during the passive rotation via the commutator apparatus and the subjects were asked to localize the sound source. Majority of the reported results were consistent with the synthesized direction suggesting that access to information provided by the vestibular organ is sufficient for the formation of the dynamic cue [42]. Several studies published since Wallach's experiments support the view that visual information is not required for the formation of the dynamic cue; see e.g.,[19], [25], [40].

In the third set of experiments, the isolated effect of visual information on the formation of the dynamic cue was investigated through visually induced illusory self-rotation using the setup shown in the figure on the left-hand side of Fig. 28. The subject was seated in the middle of a circular revolving screen patterned with black and white stripes. A loudspeaker was placed in the intersection of the subject's frontal median and azimuthal planes on the other side of the revolving screen. The screen was patterned with black and white stripes to aid in the induction of illusory self-rotation when the screen was set to revolve around the subject. Such an arrangement led to the subjects perceiving themselves to be in rotating motion rather than the screen. Because the stationary loudspeaker beyond the revolving screen remained in the

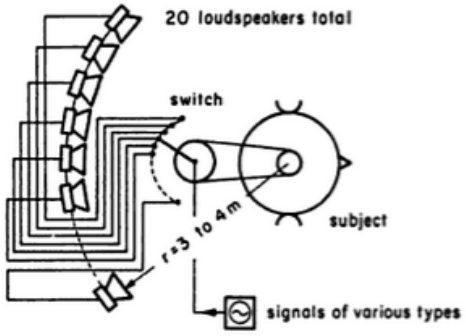


Figure 27: Experimental apparatus used in Wallach's commutator experiments. Adopted from [3].

median plane of the test subjects during their illusory self-rotation, the presented sequence of lateral angles was equivalent to that produced by an overhead sound source.

In another experiment employing the same apparatus, the loudspeaker was displaced with an angular velocity equal to that of the rotating screen but in the direction opposing the screen's rotation. This configuration is shown in the figure on the right-hand side of Fig. 28. Again, the subject perceived himself to be in rotating motion in the direction opposing that of the moving screen. Because the loudspeaker is displaced in the same direction as the subject's perceived self-rotation, the situation is analogous to that depicted previously in Fig. 25 in which a sound source was displaced along with the observer's head rotation at twice the rate of the observer's rotation. According to the principle of least displacement, the presented sequence of lateral angles should therefore cause the sound source to be perceived stationary in the hemisphere opposite to the sound source's actual location.

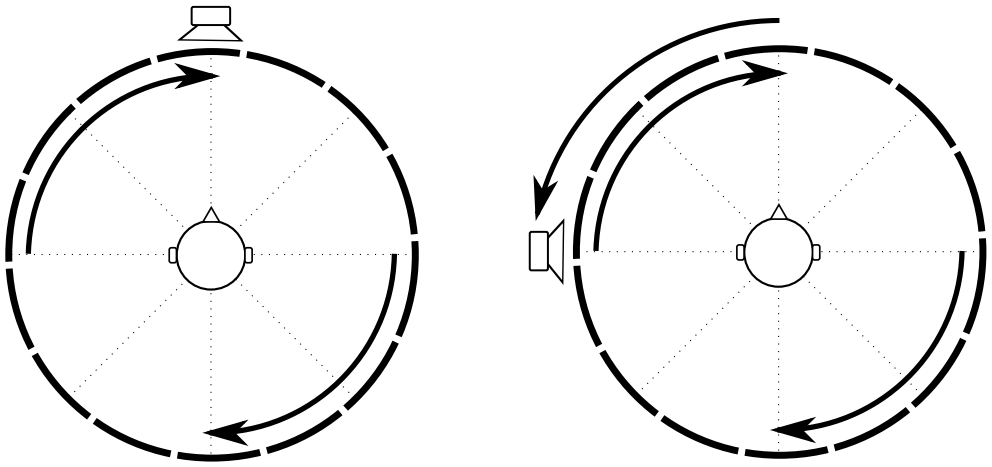


Figure 28: Experimental setup used by Wallach to test the effectiveness of visually induced illusory self-rotation in forming the dynamic cue. A striped non-transparent screen revolves around the test subject inducing an illusion of self-rotation in the direction opposing the screen's rotation. A loudspeaker is hidden beyond the screen. The figure on the left shows the situation with a stationary loudspeaker. In the figure on the right, the loudspeaker is displaced at the same angular velocity as the screen but in the opposite direction. The two cases correspond to displacement factors of one and two.

Wallach reported positive results for both of the above described vection experiments [42], suggesting that visual information is sufficient for forming the dynamic cue in the absence of vestibular and proprioceptive information. More recent experiments employing a wide-field projector to visually induce illusory self-rotation have however been unable to reproduce Wallach's results [28].

Kim et al. [19] isolated the proprioceptive cue by instructing their test subjects to keep their head stationary while seated on a motorized rotating chair undergoing slow sinusoidal oscillation covering a spatial window of 90 degrees. In order to keep their head stationary, the subjects had to counterbalance the rotation imposed by the motorized seat via the muscles of the neck. Such an arrangement minimizes the accelerating forces acting on the vestibular organ thus leaving the proprioceptors of the neck muscles as the only source providing head orientation information. The task of the subjects was to localize low-pass filtered noise stimuli presented over headphones to the frontal or rear hemiplane. The ITD-information carried by the stimuli was modified to be consistent with a situation where the head was in rotating motion and the torso remained still. The results reported in [19] suggest that the proprioceptive information furnished by the muscles of the neck is not a sufficient sensory input for the formation of accurate dynamic cues. In the light of currently available information, it therefore seems that the information provided by the vestibular organ is both required and sufficient for the salience of the dynamic auditory cue.

5 Overview of previous studies assessing cue conflict between dynamic and spectral cues

Although cue conflict between binaural cues has been widely researched, there are relatively few studies explicitly exploring cue conflict with contradicting dynamic and spectral cues. Wallach's commutator experiments presented in the previous chapter are the earliest documented studies assessing this topic. Although the conclusion reached in these publications was that spectral cues are subordinate to dynamic cues, recent studies have shown that this claim does not hold under all stimulus conditions. This chapter provides an overview of the handful of studies that have assessed the topic of sound localization under conflicting dynamic and spectral cues since Wallach's early experiments. Although the experimental setups employed in these studies differ greatly from each other, their conclusions are in general agreement. For the purposes of this overview, the studies have been divided into two groups based on whether the cue conflict targeted vertical or horizontal localization.

5.1 Vertical localization

In Section 4.4 it was shown that the absolute value of source elevation can be resolved via dynamic cues through monitoring the rate of change of the lateral angle with respect to the displacement rate of the interaural axis in the horizontal plane. Furthermore, hemisphere ambiguity is resolved by the direction of change in the binaural cues as a result of side to side pivoting motions of the head that displace the interaural axis in the vertical dimension. The direction of change in the binaural cues—and thus the apparent source location indicated by the dynamic cues—can be reversed by interchanging the left and right ear signals via an appropriate apparatus. A device that performs such a transposition is generally referred to as a pseudophone [46].

In order to set the dynamic cues in opposition to the spectral cues, Martens et al. [27] made use of what the authors call a 'binaural hearing instrument' that allowed the acoustic signals present at the two ears to be interchanged without significant spectral distortions. The utilized pseudophone was essentially a pair of electret microphones integrated into a pair of isolating earphones. When in use, the microphone capsules rested at the entrance of the subject's ear canals, thus capturing the spectral cues formed by shoulder reflections and the reflections formed in the cavities of the pinnae.

In the test scenario, subjects were fitted with the binaural hearing instrument and instructed to walk past a vertical loudspeaker array consisting of loudspeakers positioned both above and below the subject's ear level. In each trial, one loudspeaker in the array radiated a continuous speech signal as the subject walked towards it. Once the subject had passed the loudspeaker array, his/her task was to answer whether he/she perceived the test signal above or below ear level. Each subject completed several trials under both normal and reversed binaural conditions.

The results of this experiment are shown on the right-hand side of Fig. 29. The dashed line indicates that the majority of trials under interchanged binaural conditions

resulted in hemisphere reversals. The pseudophone apparently retained the spectral cues necessary for hemisphere discrimination under both binaural conditions, as indicated by the results shown on the left-hand side of Fig. 29 obtained from a control experiment where the subjects remained stationary. The fact that the elevation judgements were reversed under interchanged binaural conditions only in the presence of dynamic cues, was taken by the authors as evidence that the erroneous dynamic cues overrode the apparently accurate spectral cues to source elevation.

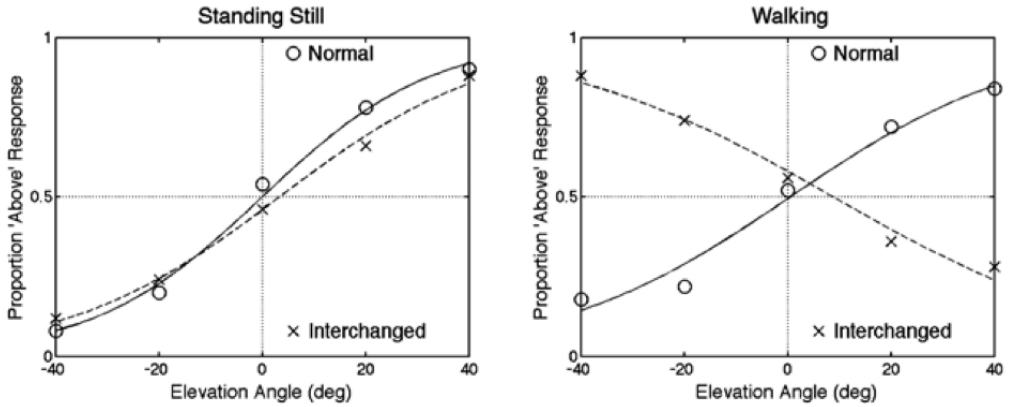


Figure 29: Experimental results from the binaural hearing instrument experiment with speech stimuli as reported in [27]. An electroacoustic interchange of the acoustic signals between the left and right ears resulted in reversal of elevation judgements of speech stimuli presented during walking. Stimuli presented below the interaural axis were localized above the said axis and vice versa.

In a continuation study by Martens et al.[26], the same experimental setup was employed using 20 ms bursts of white noise pulsed at a rate of 10 Hz as the stimulus. The results for trials where the source was located above ear level are shown in Fig. 30. The figure on the left-hand side shows the results for one listener who was able to localize the sound source accurately under normal binaural conditions. When presented with interchanged ear signals however, the localization accuracy of this observer dropped to just above 50% for the standing condition and decreased further in the walking condition. The mean results averaged over ten participants however showed no significant differences between the results obtained with interchanged ear signals when walking or standing.

Based on these results, Martens et al. suggest, that conflicting spectral and dynamic cues lead to an unstable localization process where the subjects switch between the cues they attend to in making the localization judgement [26]. Furthermore, cue preference seems to vary between subjects as implied by the results shown in Fig. 30.

5.2 Horizontal localization

The auditory system's tendency to prefer stationary source interpretations over dynamic ones was introduced in Section 4.4.1. Furthermore, it was shown that apparent source elevations can be synthesized by moving a sound source along with the observer's head movements to produce a synthetic sequence of lateral angles corresponding to a stationary phantom source. The magnitude of the resulting

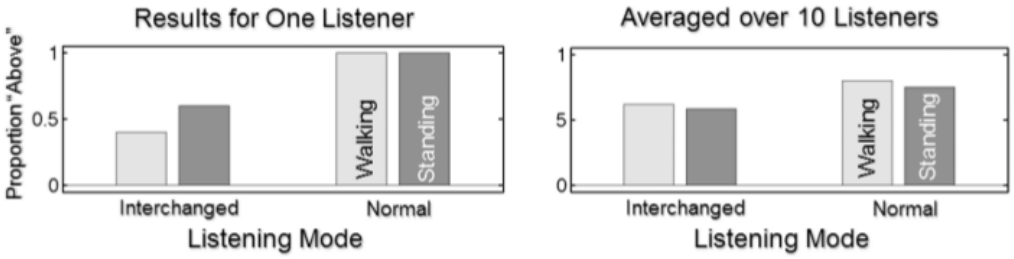


Figure 30: Experimental results from the binaural hearing instrument experiment with pulsed white noise stimuli as reported in [26]. An electroacoustic interchange of the acoustic signals between the left and right ears resulted in inconsistent reversal of elevation judgements of pulsed white noise stimuli presented during walking.

elevation angle is a function of the displacement factor, i.e., the ratio between the angular displacement of the sound source for a given angular displacement of the interaural axis. The majority of the localization studies reviewed in this section employed the principle of least displacement to induce hemiplane reversals via the dynamic cues, while the spectral information indicated a source in the opposite hemiplane, thus creating a discrepancy between the localization cues.

Macpherson used a virtual auditory space recreation of Wallach's commutator setup shown previously in Fig. 27. Subjects were fitted with a trackable headband that allowed stimulus signals presented over headphones to be dynamically processed according to the subject's instantaneous head orientation. In an early pilot experiment [23], four subjects were presented with head-motion-coupled stimuli of various bandwidths as they performed 40-degree head rotations. The four employed stimulus groups provided different combinations of binaural and spectral cues. A wideband noise band of 0.5 - 16 kHz and a high-pass noise band of 4 - 16 kHz provided access to accurate spectral cues with or without signal energy in the ITD-region. Low- and high-frequency narrow band noises of 0.5 - 1 kHz and 6.0 - 6.5 kHz provided ITD or ILD information without access to accurate spectral cues. A displacement factor of two was used to move the stimuli in virtual auditory space implemented with individually measured HRTFs. The task of the subjects was to report whether they perceived the virtual sound source as moving or stationary; a stationary interpretation was taken as evidence for the dominance of dynamic cues over spectral cues and vice versa.

The percentage of trials where the source appeared to be moving was 43% for the wideband noise and increased to 58% for the high-pass noise. The corresponding percentages for the low- and high-frequency narrowband noises were 14% and 3% respectively. The results for the wideband and high-pass noises seem to indicate that in the presence of conflicting dynamic and spectral cues, neither cue dominates on average. The increased motion rate in the case of the high-pass noise implies that cue preference shifts to spectral cues as low-frequency energy is removed from the stimulus. The low motion rates of the narrowband stimuli imply that the static, least displacement interpretation of dynamic cues is preferred in the absence of spectral cues for low- and high-frequency narrowband stimuli.[23]

In another experiment employing the same test apparatus [24], MacPherson

presented subjects with low-pass filtered noise stimuli of different attenuation slopes as depicted on the left-hand side of Fig. 31. The task of the subjects was to indicate whether they perceived the virtual source in the frontal or rear hemiplane. The results are shown on the right-hand side of Fig. 31. The blue bars show the results of a control experiment performed under static head conditions to test the accuracy of the individually measured HRTFs. The results for the head movement condition indicate, that the least displacement interpretation of dynamic cues dominated the localization of the low-frequency narrowband noise stimulus. Although the author reports intersubject variability in the results corresponding to the low-pass filtered stimuli, the average performance was just below 50% for the five low-pass filtered spectra with varying amount of high-frequency energy.

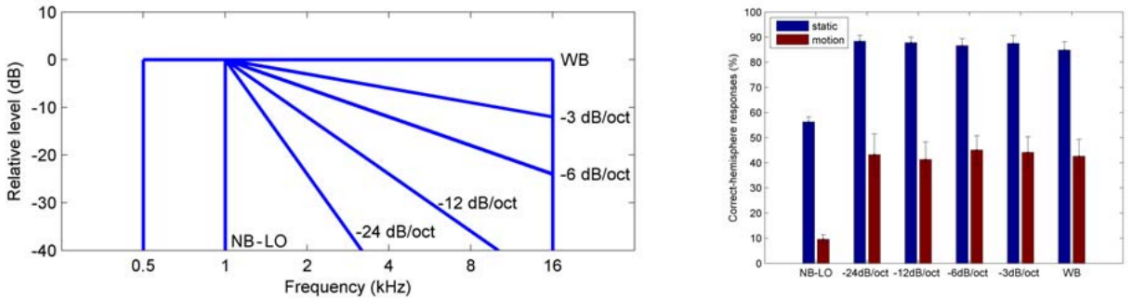


Figure 31: Stimulus spectra and the corresponding localization results as reported in [24]. A displacement factor of two resulted in subjects correctly identifying the correct hemisphere of stimulus presentation in only 10% of the trials when a narrowband low-frequency noise stimulus was used. The localization rate approached 50% when the for wideband stimuli of various high-frequency cutoff slopes.

Brimijoin and Akeroyd conducted a similar experiment using loudspeakers [4]. Subjects were seated in the middle of a circular loudspeaker array and fitted with a head tracking device. In the test scenario, subjects rotated their heads through spatial windows of 30 degrees. Displacement factor of two was used to move a low-pass filtered speech stimulus between loudspeakers in the circular array by using equal-power panning. For a given trial, stimuli could be presented from loudspeakers in either the frontal or rear hemiplanes. The task of the subjects was to report the hemiplane where they localized the stimulus.

The averaged results over seven participants shown in Fig. 32 indicate, that hemiplane reversal due to the least displacement interpretation of dynamic cues was most often reported for speech stimuli low-pass filtered at 500 Hz. The average reported rate of hemiplane reversal approached 50% as the cutoff frequency of the stimuli increased.

In a related experiment utilizing the same head tracking equipment and circular loudspeaker array, Brimijoin et al. investigated the effect of head movements on the internalization of stimuli presented over loudspeakers. Speech stimuli were presented from locations on the azimuthal plane and coupled to the observer's head movements using equal power panning and a displacement factor of one. The subject's task was to rotate his/her head about ±15 degrees and report whether a given stimulus was localized inside the head—i.e., internalized—or not. The results shown in Fig. 33

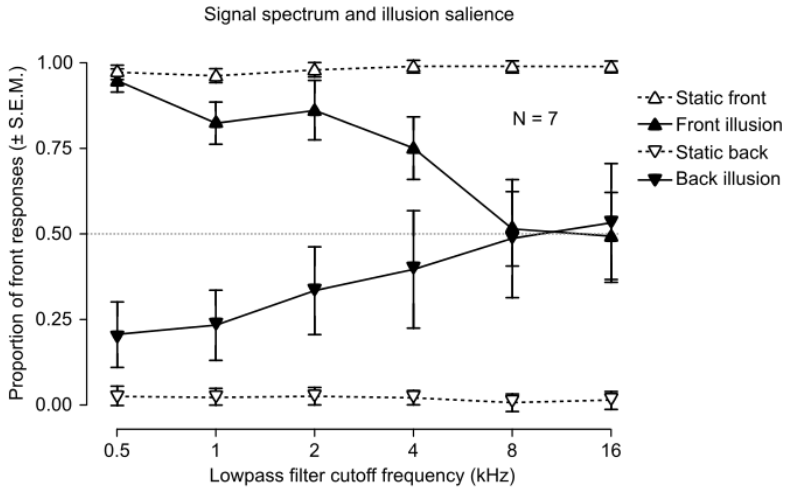


Figure 32: Experimental results from the localization experiment described in [4]. Open triangles correspond to control trials where the sound source was static, i.e., the displacement factor was zero. Erroneous dynamic cues formed by a displacement factor of two dominated localization for speech stimuli presented from either hemiplane when the stimuli were low-pass filtered at 500 Hz. When high-frequency components were added to the stimuli, localization accuracy approached chance rate.

indicate, that the head-motion-coupled panning resulted in a an increased proportion of internalized trials with respect to the control trials where the stimulus remained stationary during head movements.

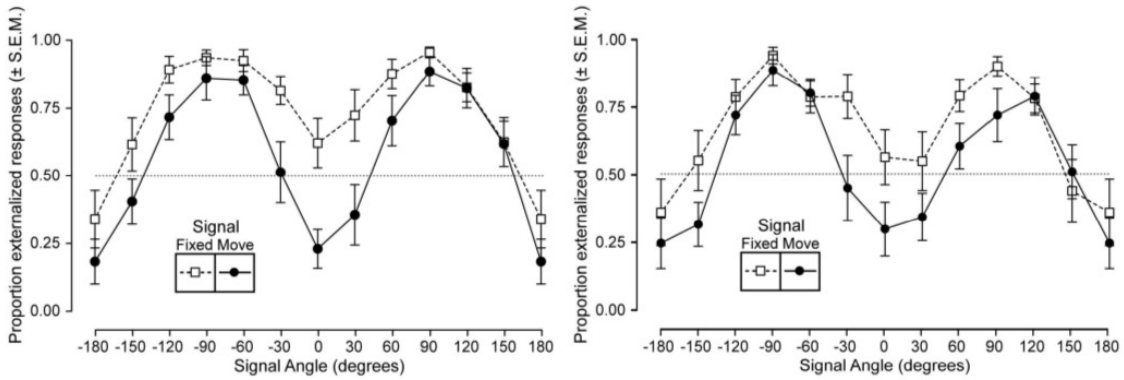


Figure 33: Internalization of head-motion-coupled stimuli as a function of azimuth angle. Black and white squares denote results from trials using a displacement factors of one and zero respectively. The figure on the left-hand side shows the results for unfiltered speech stimuli. The right-hand side shows the results obtained with speech stimuli low-pass filtered at 500 Hz. Internalization was most common for stimuli presented from the median plane. Adopted from [5].

The experiment conducted by Kawaura et al. differs from the other studies reviewed in this section as their approach was not based on the principle of least displacement. In an experiment described in [18], a musical stimulus presented over headphones was filtered with static HRTFs corresponding to source locations on the azimuthal plane. A head-motion-coupled dynamic ITD cue corresponding to a source location on the hemiplane opposite to that implied by the applied HRTF was added to the stimulus to create the cue conflict. In the experimental scenario, subjects

rotated their heads in the horizontal plane and reported the perceived source location. Two of the three subjects localized the virtual source to the direction consistent with the dynamic ITD cues, while the third subject localized the source consistently to the rear hemiplane, regardless of the location implied by the dynamic ITD [18].

Based on the studies reviewed in this section, it seems that dynamic and spectral cues are weighted equally leading neither cue to systematically dominate localization if the stimulus provides access to both cues. Low-pass filtered stimuli that do not provide access to spectral cues seem to be localized according to the principle of least displacement. Similarly, narrowband stimuli appear to be localized according to dynamic cues regardless of center frequency. In addition to the source directions implied by the principle of least displacement, erroneous dynamic cues produced by using a head-motion-coupled panning arrangement with a displacement factor of one can result in the internalization of stimuli presented via loudspeakers under free-field conditions. At least one case has been reported where a subject localized virtual sources systematically to the rear hemiplane when presented with stimuli with conflicting dynamic and spectral cues.

6 The conducted listening test

A common feature to many of the experiments reviewed in the previous chapter is the utilization of test stimuli with poorly defined spectral characteristics. While musical and speech stimuli typically contain both low- and high-frequency energy, their overall spectral content tends to be modulated over time and biased towards low frequencies, thus providing unbalanced stimulation of the auditory filters. Similarly, a white noise stimulus with constant power spectral density per Hz provides more signal energy to the wider filters at high center frequencies. In order for the relative weighting of these two cues to be reliably assessed, the stimulus spectrum should be well defined and provide equal stimulation of the auditory filters.

This chapter describes the implementation and results of a listening test, where auditory localization under conflicting dynamic and spectral cues was studied with a filtered pink noise stimulus. The power spectral density of pink noise results in stable and approximately equal stimulation of the auditory filters across the audio band. Subjects were asked to localize burst sequences containing various ratios of signal energy in the low-frequency and spectral cue regions.

The principle of least displacement was used to set the dynamic cues in conflict with the spectral cues. Misleading dynamic cues were presented to the subjects via a headtracking apparatus that allowed acoustic stimuli to be presented over loudspeakers from a direction determined by the instantaneous orientation of the subject's head. In the experimental scenario, subjects performed horizontal rotations during which bursts of pink noise were presented from unelevated source locations on the subject's median plane. The headtracking procedure produced static binaural signals corresponding to those produced by an overhead source.

The task of the subjects was to report the perceived direction of the burst sequence via a two-alternative forced choice procedure. One answer alternative represented the actual direction where the burst sequence was presented; the other alternative corresponded to all other localization judgements. Answers that corresponded with the actual location of the burst sequence were interpreted as accurate spectral cues dominating the localization judgement in that trial. Conversely, answers that did not correspond with the actual location of the stimuli were interpreted to be a result of the misleading dynamic cues dominating localization judgement.

The test hypothesis was that localization accuracy is a function of stimulus spectrum. Low-frequency components in the ITD-region provided access to misleading dynamic cues and high-frequency components provided access to spectral cues that revealed the actual direction of the burst sequence. Stimuli containing low-frequency components were therefore expected to require more high-frequency components for accurate localization than high-pass filtered stimuli lacking low-frequency energy. The high-frequency content of the stimuli was adjusted in a psychoacoustically informed way by filtering the pink noise stimuli in one unit increments on the ERB-rate scale.

The following sections describe the implementation of the experiment in more detail. Experimental results and their statistical analysis are presented in Section 6.5.

6.1 Participants

Fourteen subjects reporting normal hearing participated in the conducted listening tests. All subjects were staff or students at the Department of Signal Processing and Acoustics at Aalto University.

6.2 Stimuli

The stimuli were 10 ms bursts of filtered pink noise gated with 1 ms raised cosine ramps. The stimuli were divided into four groups according to the applied filtering scheme. Each group is characterized by three parameters: low corner frequency and the upper and lower limits of stimulus high cutoff frequency. Low corner frequency denotes the fixed -3 dB low cutoff frequency common to the filters used to create the samples within any one group. The lower and upper limits define the span of filter -3 dB high cutoff frequencies used to create the stimuli within the group. The lower limits of each group were chosen according to the limitations of the filter synthesis algorithm used to create the FIR-filters. The high cutoff frequency was increased in increments of 1 z_{ERB} from the lower limit to the upper limit and a batch of 30 stimuli was created for each cutoff frequency pair. The stimulus group parameters are collected in Table 3 below. The burst sequence used in each trial was formed from seven randomly chosen unique samples from the appropriate batch to ensure that the correlation between bursts in the sequence was sufficiently low.

Group	Low corner (z_{ERB})	Lower limit (z_{ERB})	Upper limit (z_{ERB})
1	0	7	41
2	10	16	41
3	15	17	41
4	21	22	41
Group	Low corner (Hz)	Lower limit (Hz)	Upper limit (Hz)
1	0	259	19000
2	445	1061	19000
3	929	1208	19000
4	1985	2238	19000

Table 3: Description of the stimulus groups used in the experiment. The column labeled 'low corner' denotes the low cutoff frequency of the filters applied to the samples within a given group. The lower and upper limit columns indicate the lowest and highest filter high cutoff frequencies applied to samples within a given group.

The samples were synthesized digitally by first creating a 10 second vector of white noise using a random number generator. The spectrum was then colored pink by weighing the white noise vector's amplitude spectrum by a factor of $1/\sqrt{f}$; this corresponds to a $1/f$ decay characteristic in the power spectral density. The resulting pink noise vector was then filtered with a 5th order FIR butterworth filter with corner frequencies appropriate to the sample under synthesis. The RMS-value of the filtered noise vector was normalized to reduce loudness variations between samples with

varying frequency content. Thirty non-overlapping 10 ms sections were extracted from the filter's steady-state output vector and 1 ms raised cosine ramps were then applied to the onset and offset of each sample to avoid the spectral spreading associated with sudden transients. All stimuli were synthesized and presented at a sample rate of 48 kHz. The effect of finite FIR-filter transition-bands and stop-band attenuation was reduced by setting the A-weighted noise floor within the test chamber 20 dB below the stimulus level. The noise floor was raised by presenting low-amplitude, uncorrelated pink noise from six loudspeakers within the test chamber.

6.3 Experimental setup

The experiment was conducted in an anechoic chamber fitted with a set of six motion tracking infrared cameras and a multichannel loudspeaker system. The test subjects were fitted with a reflective headband that allowed the camera system to track the orientation of the subject's head via triangulation at a rate of 100 Hz; linear interpolation was used to update the head orientation coordinates between samples. The head tracking interface was running on a dedicated computer that delivered head orientation data to another computer that controlled the progression of the test. The second computer was used to route the test stimuli to the 16 loudspeakers via a multichannel digital-to-analog converter. The subjects reported their answers via a tablet computer with a wireless connection to the test computer. The signal chain employed in the experiment is shown in Fig. 34.

The loudspeaker arrangement used in the experiment is illustrated in Fig. 35. The stimuli were presented from two symmetric loudspeaker clusters suspended on a circular rail of 2.1 m radius. The angular separation of the seven adjacent loudspeakers in both clusters was 5 degrees as measured from the middle point of the loudspeaker array. The six loudspeakers located at ±15,±90 and ±165 degrees radiated uncorrelated pink noise to raise the semi-diffuse noise floor within the test chamber.

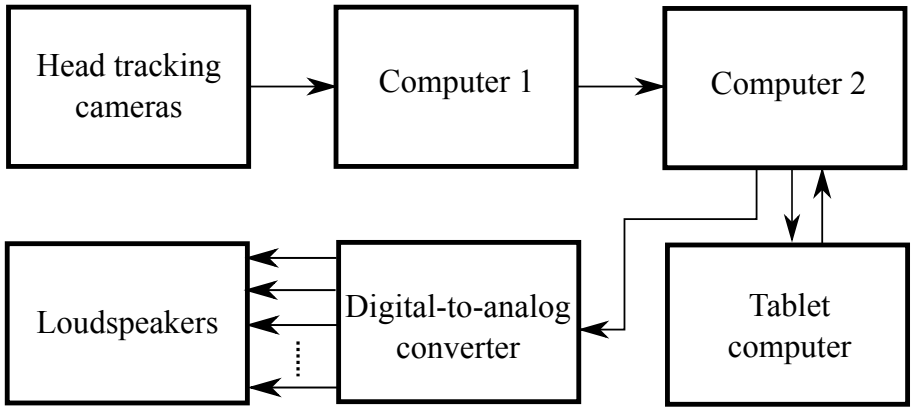


Figure 34: Signal chain used in the experiment.

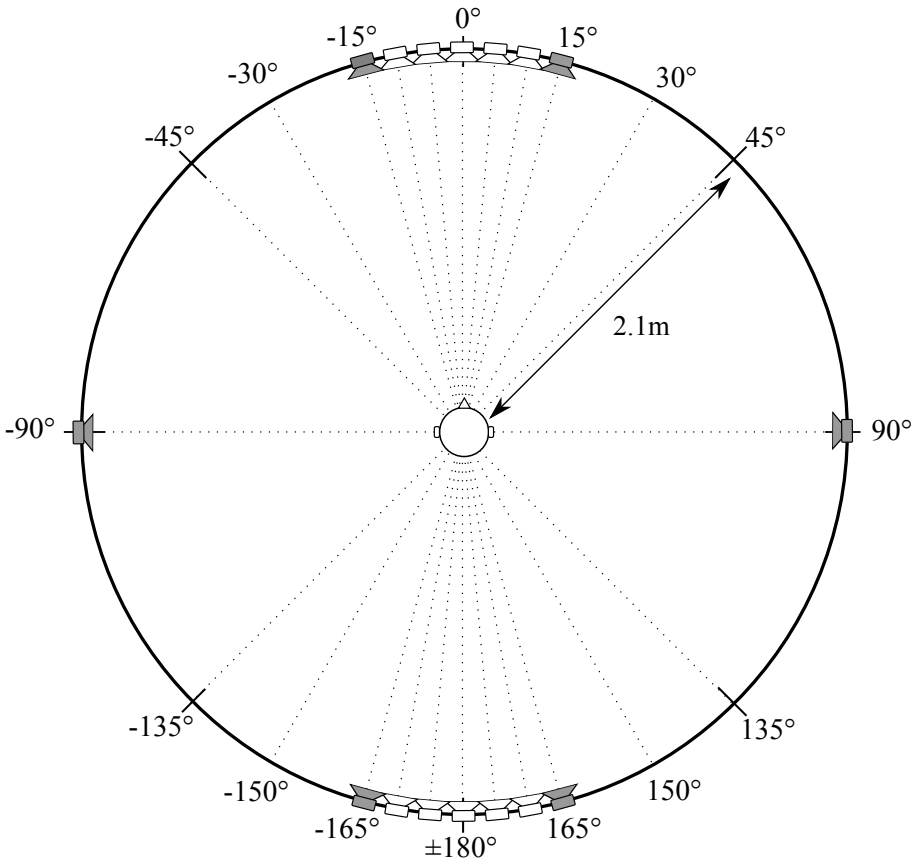


Figure 35: Loudspeaker configuration used in the listening test. The six loudspeakers used to raise the noise floor within the chamber are marked gray. Stimuli were randomly assigned to the front and rear clusters from trial to trial.

6.4 Test procedure

The conducted experiment consisted of two separate sessions testing localization performance in the frontal and rear hemiplanes. All fourteen subjects participated in both hemiplane tests. Half of the subjects were randomly chosen to participate in the rear hemiplane test before the frontal hemiplane test; the remaining half were assigned to start with the frontal hemiplane test.

In the experimental scenario, test subjects were seated on a rotating chair in the middle of the loudspeaker array shown in Fig. 35. In each trial, the subject first rotated the chair to orient themselves to approximately 30 degrees to either side of the middle loudspeaker of the front cluster. Correct orientation was signalled to the subject by a green light on the tablet computer once the head tracker data showed that the subject was oriented appropriately. A single trial could then be initiated via the tablet computer. During the trial, subjects rotated towards the other side of the loudspeaker cluster at a rate of approximately 50 degrees/second. Head orientation data was used to produce a displacement factor of one by triggering noise bursts from individual loudspeakers as they coincided within ±1 azimuthal degrees of the subject's median plane. The task of the subjects was to report whether they perceived the sequence of bursts in the vicinity of the loudspeaker cluster in the tested direction

or somewhere else via a two-alternative forced choice procedure. The subjects had the possibility to retry each trial by repeating the above described steps before giving an answer. All subjects were familiarized with the experimental procedure in a supervised training session before starting the experiment. No feedback was given to the subjects regarding their performance in the experiment.

In both hemiplane experiments, trials were randomly assigned to either front and rear loudspeaker cluster from trial to trial. Burst sequences that were not presented from the cluster in the tested direction (e.g., burst sequences from the rear cluster in the frontal plane test) acted as dummy trials and did not influence the progress of the test. All subjects were familiarized with the experimental procedure in a supervised training session before starting the experiment.

The test progressed according to an adaptive staircase procedure with the four sample groups assigned to individual tracks. The adaptive procedure altered the high cutoff frequency of the presented stimuli in one z_{ERB} increments according to a three-down, one-up rule. The initial upper cutoff frequency of each track was set to 34 z_{ERB} corresponding to approximately 8.8 kHz; all tracks therefore contained frequencies in the spectral cue region at the beginning of the test. The sequence of tested tracks was chosen randomly between all unfinished tracks excluding the previously tested track. In situations where only one track remained unfinished, dummy trials from randomly chosen finished tracks were presented between trials of the unfinished track.

The experiment finished when all of the four adaptive tracks had terminated. A track could terminate in one of three ways: by crossing the upper or lower limits of the track (see Table 3 for track limits) or by converging. A total of ten reversals was required for a track to terminate by convergence.

6.5 Results

Experimental results from the frontal and rear hemiplane experiments for fourteen subjects are presented in Figures 36 and 37 respectively. Trial number is shown on the x-axis; the y-axis denotes the high cutoff frequency of the stimulus corresponding to a given trial expressed in equivalent rectangular bandwidths. The four tracks correspond to the four sample groups with varying low-frequency content described in described previously in Table 3.

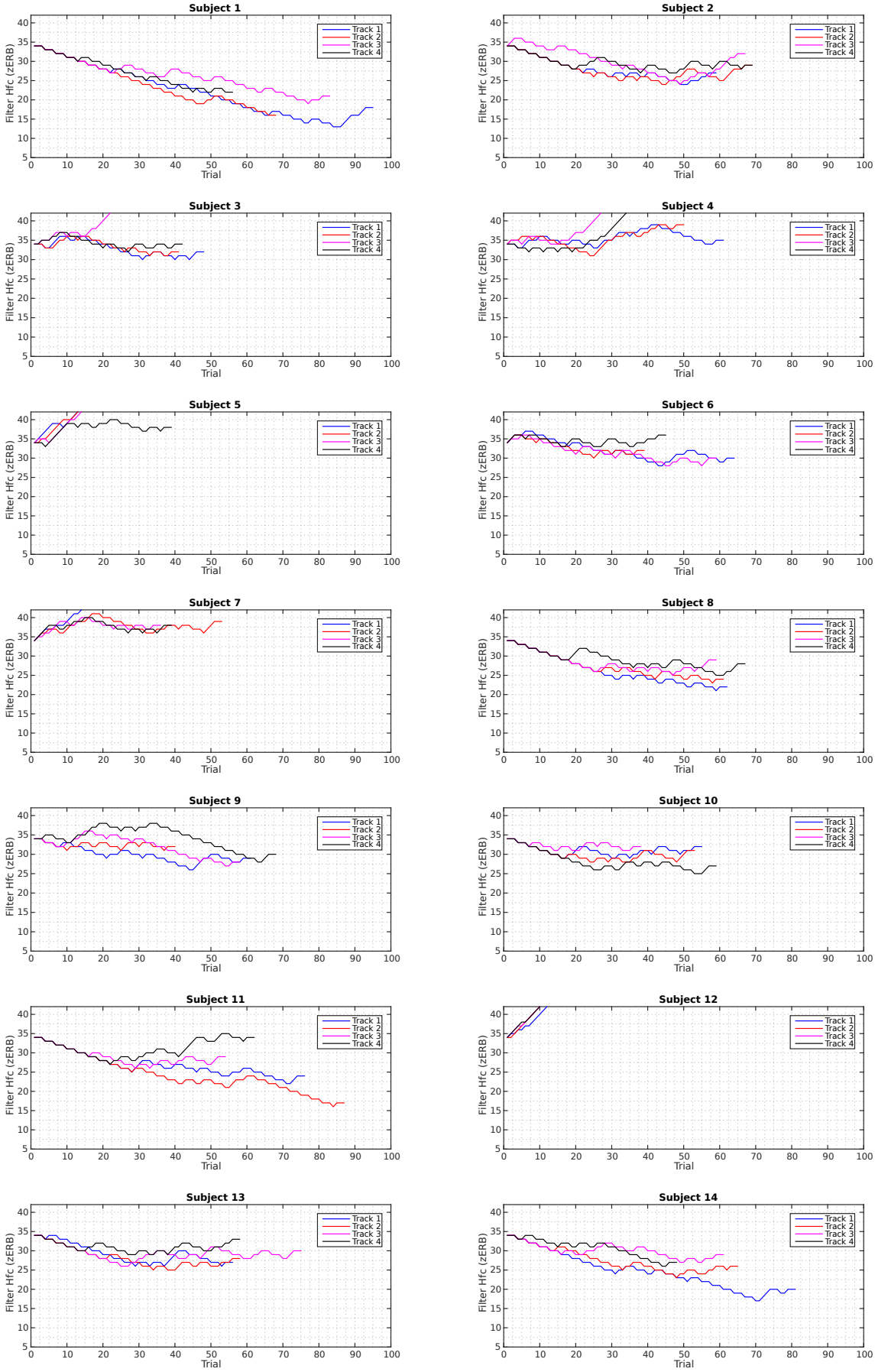


Figure 36: Frontal hemiplane test results.

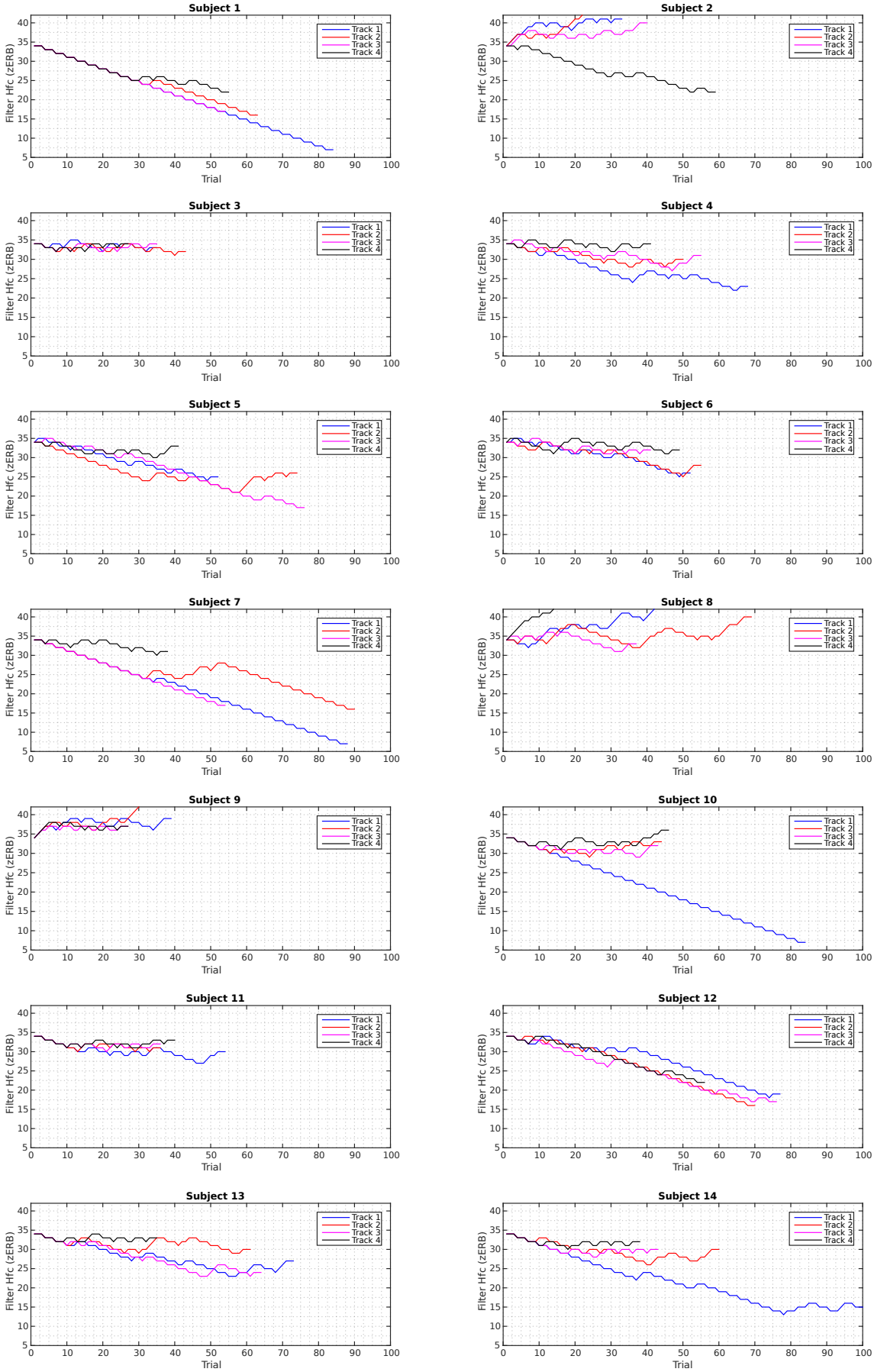


Figure 37: Rear hemiplane test results.

6.5.1 Statistical analysis of convergent tracks

In this section, the statistical analysis methods introduced in Section 3.3 are applied to the obtained results to investigate the effect of low-frequency signal energy to the mean reversal value of the convergent tracks. While the majority of the tracks terminated by completing the required ten reversals, it is clear that a number of them did not converge to a value representing a threshold. These tracks are disqualified from the analysis by applying a condition that the last five reversals of the tracks have to be within five ERBs for a track to qualify as convergent. Examples of tracks that were disqualified based on this condition are shown in Fig. 38.

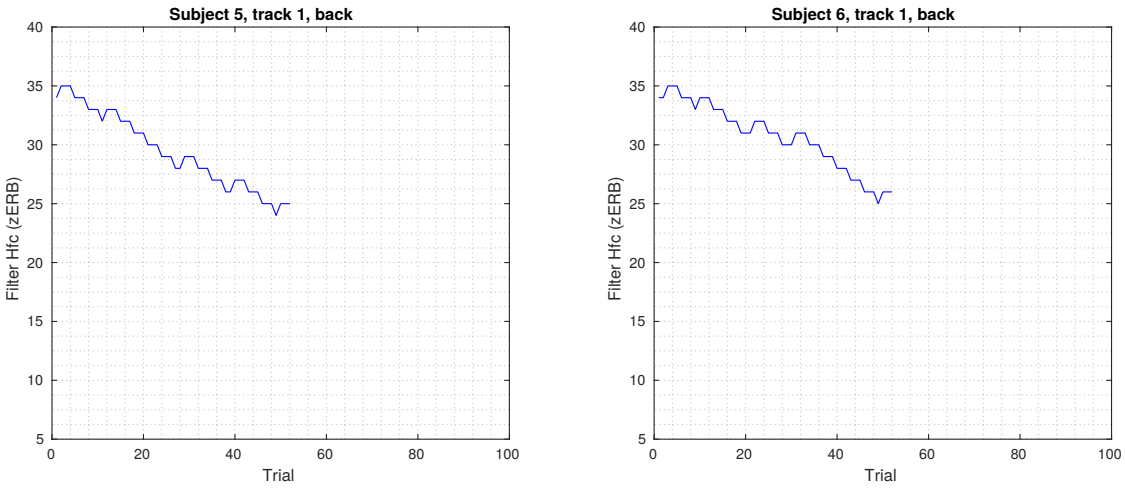


Figure 38: Examples of tracks that did not fulfill the applied convergence condition.

Summarized track termination data is shown in Table 4. Cells marked 'high' and 'low' indicate that the track terminated by crossing the upper or lower limits and did not complete the required ten reversals. Numerical values indicate the mean z_{ERB} value of the last five reversals for tracks that terminated by convergence. Cells marked with an asterisk (*) denote that the track did not fulfill the convergence condition and are therefore excluded from the statistical analysis. The remaining numerical threshold values were analysed according to the techniques outlined in Section 3.3. Summary statistics and box plots of the data fulfilling the convergence criterion are shown in Table 5 and Fig. 39 for both hemiplane tests.

In the case of the frontal plane results, a Shapiro-Wilk normality test yields significance values of 0.765, 0.473, 0.103 and 0.518 for the means of the observations for tracks 1, 2, 3 and 4 respectively. The null-hypothesis is therefore not rejected and the assumption of normality is valid for all four sample groups at the applied 5% test level. Levene's test for homogeneity of variances gives a result of 0.724 thus indicating that the assumption of homogeneous variances is not violated at 5% test level. The data therefore conforms to the assumptions of ANOVA and the standard one-way analysis can be carried out without modification. The analysis yields a significance value of 0.186 indicating no statistically significant differences between the means of the sample groups. The ANOVA table for the front test is shown at the top of Table 6.

	Front				Back			
Subject	Track 1	Track 2	Track 3	Track 4	Track 1	Track 2	Track 3	Track 4
1	15.4	Low	*22.2	Low	Low	Low	Low	Low
2	26.2	26.4	*27.6	29.0	40.6	High	37.6	Low
3	31.0	31.8	High	33.6	33.2	32.4	33.2	33.6
4	36.2	*37.8	High	High	24.4	29.2	30.4	33.4
5	High	High	High	38.0	*26.2	25.2	Low	31.6
6	30.2	31.6	29.6	34.4	28.8	*29.6	31.6	32.6
7	High	37.6	37.6	36.8	Low	Low	Low	31.6
8	22.4	24.4	26.8	27.4	High	*36.0	34.2	High
9	28.6	32.2	*29.2	*33.8	38.0	High	36.6	36.6
10	31.0	29.4	32.2	27.0	Low	32.2	30.6	33.2
11	24.4	*20.2	28.0	*33.8	29.4	30.8	31.6	32.2
12	High	High	High	High	*25.8	Low	Low	Low
13	27.2	26.8	29.4	30.8	25.4	31.2	24.8	32.6
14	*22.0	25.2	28.4	*29.6	15.4	28.4	29.6	31.6
High #	3	2	4	2	1	2	0	1
Low #	0	1	0	1	3	3	4	3
Converged #	10	9	7	8	7	7	10	10

Table 4: Track termination data.

Front							
Track	Mean	Median	Variance	Std. Deviation	Min	Max	Range
1	27.3	27.9	32.4	5.7	15.4	36.2	20.8
2	29.5	29.4	18.0	4.2	24.4	37.6	13.2
3	30.3	29.4	13.2	3.6	26.8	37.6	10.8
4	32.1	32.2	17.7	4.2	27.0	38.0	11
Back							
Track	Mean	Median	Variance	Std. Deviation	Min	Max	Range
1	29.5	29.4	75.1	8.7	15.4	40.6	25.2
2	29.9	30.8	6.5	2.5	25.5	32.4	7.2
3	32.0	31.6	13.5	3.7	24.8	37.6	12.8
4	32.9	32.6	2.3	1.5	31.6	36.6	5.0

Table 5: Summary statistics for mean reversal data.

Following the same procedure with the data from the rear hemiplane results, a Shapiro-Wilk's normality test yields significance values of 0.907, 0.341, 0.769 and 0.016 for tracks 1,2,3 and 4 respectively, indicating normality in three of the four sample sets. Levene's test produces a statistically significant value of 0.001, indicating that the data violates the assumption of homogeneity of variances; the robust significance value is therefore obtained via a Welch-ANOVA rather than a standard one-way

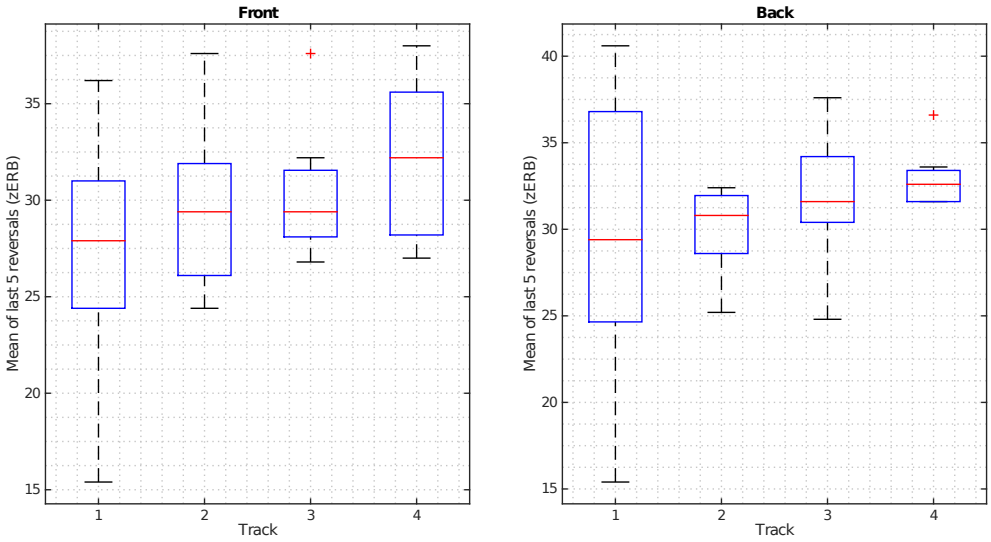


Figure 39: Box plots of mean reversal data computed from tracks fulfilling the convergence criterion. The red crosses denote suspected outliers but were included in the analysis.

ANOVA, resulting in a significance value of 0.097. The null hypothesis is therefore accepted at 5% test level and no statistically significant differences can be claimed between the group means. The ANOVA table for the back sector results is shown at the bottom of Table 6; the corresponding Welch-ANOVA results are shown in Table 7.

Front					
	Sum of squares	Degrees of freedom	Mean Square	F-ratio	Significance
Between-groups	109.15	3	36.38	1.71	0.186
Within-groups	639.12	30	21.30		
Total	748.27	33			
Back					
	Sum of squares	Degrees of freedom	Mean Square	F-ratio	Significance
Between-groups	67.25	3	22.42	1.07	0.378
Within-groups	631.15	30	21.04		
Total	698.40	33			

Table 6: ANOVA tables for mean reversal data.

Back				
	Statistic	df1	df2	Significance
Welch	2.584	3	13.39	0.097

Table 7: Welch-ANOVA for the rear hemiplane data.

7 Discussion

This section starts with a discussion of the statistical analysis results presented in the previous section and continues by outlining some general remarks given by the test subjects in informal discussions that took place after the experiment. The topics of cue weighting and differences in localization performance between frontal and rear hemiplanes are then discussed based on the experimental results. The section ends with a discussion of methodological issues and their possible effects on the obtained results.

7.1 Statistical analysis

The ANOVA-based statistical analysis performed in Section 6.5.1 is fundamentally limited by the fact that a large number of the measured adaptive tracks did not converge to a value representing a meaningful threshold. The statistical analysis had therefore to be carried out with a reduced dataset that does not represent the data as a whole. Moreover, the fact that the four sample groups had non-identical lower limits for the high cutoff frequency further complicates the analysis. Track 1 for example could converge as low as 7 z_{ERB} while track 4 could never converge below 22 z_{ERB}, thus unbalancing any numerical comparison of the convergence values. Because of these reasons, one should be cautious in interpreting the obtained analysis results, which in this case have to be deemed inconclusive.

7.2 General remarks made by the test subjects

In general, the majority of the subjects described the localization task as difficult. Some subjects characterised the auditory image formed by the burst sequences as diffuse and difficult to localize to any one direction. A number of subjects experienced internalization of the free-field stimuli, resulting in the auditory image being localized inside the subject's head; such reports are in agreement with the internalization study of Brimijoin et al. reviewed in Section 5.2. Another commonly reported experience was that the stimuli were localized to diagonal locations above 45 degrees of elevation. It is assumed that these percepts did not significantly alter the outcome of the adaptive tracks as the 'elsewhere' option of the utilized two-alternative forced choice answering scheme essentially covered all of these source locations.

One subject localized the last burst of each sequence to the rear hemiplane regardless of the apparent location of the burst sequence as a whole. The subject reportedly ignored this reversal tendency in the localization task and gave his answers according to the apparent source location as perceived before the last burst.

7.3 Cue weighting

Although the obtained results vary significantly between subjects, some general trends can be observed in the obtained results. It appears that some subjects gave greater weight to the misleading dynamic cues even when the stimuli provided access

to accurate spectral cues, thus leading to poor overall localization performance. On the other hand, some subjects were able to localize narrowband burst sequences accurately to the correct hemiplane even in the absence of accurate spectral cues. A considerable number of subjects displayed inconsistent localization performance, suggesting that these subjects switched between the cue they attended to in making the localization judgement throughout the duration of the experiment.

Examples of poorly localized tracks can be seen for example, in the frontal plane results for subjects 4,5,7, and 12 and rear plane results for subjects 2,8, and 9. The results of these subjects contain a significant number of tracks that terminated by crossing the upper limit or converged to a high stimulus cutoff frequency. Two example tracks are plotted in Fig. 40 for illustration purposes. Since the stimuli presented to the subjects in these tracks were wideband and therefore provided accurate spectral cues, these subjects apparently systematically based their localization judgements on the misleading dynamic cues regardless of stimulus bandwidth.

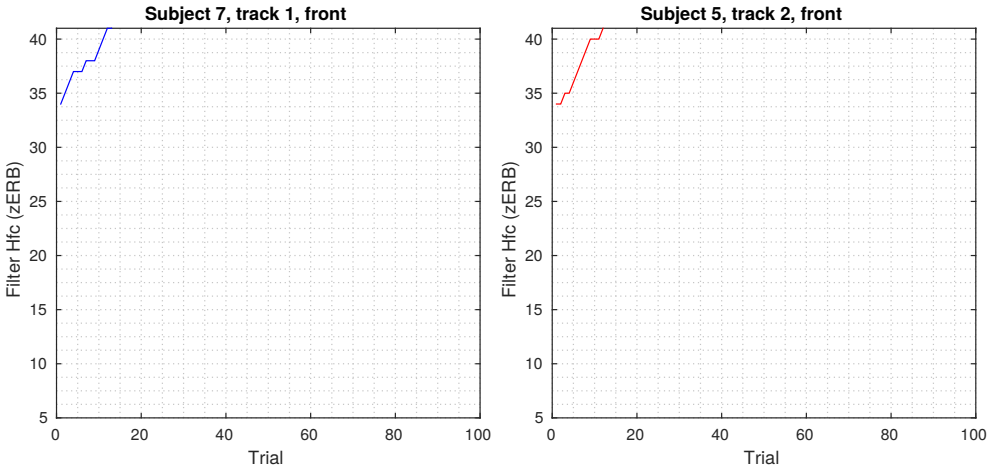


Figure 40: Examples of tracks displaying poor localization performance despite wide stimulus bandwidths.

Another group of subjects were able to localize narrowband burst sequences accurately to the correct hemiplane even in the absence of accurate spectral cues. Examples of tracks showing consistently accurate localization performance can be seen in the frontal plane results for subjects 1 and 14 and the rear hemiplane results for subjects 1,7,10,12, and 14. The results obtained from these subjects include a significant number of tracks that terminated by crossing their lower limits or converged to a low stimulus cutoff frequency. Fig. 41 illustrates two examples of tracks that were localized accurately in a consistent manner. The fact that these subjects were able to accurately localize narrowband stimuli in the absence of spectral cues suggests that the duration of the individual bursts in the stimulus sequences was sufficiently long for these subjects to extract accurate dynamic cues from the individual bursts, rather than interpreting the dynamic cues formed by the burst sequence as a whole. If some subjects did indeed localize the stimuli in this manner, it cannot be said that the experimental design was able to correctly separate localization judgements based on spectral cues from those based on dynamic cues.

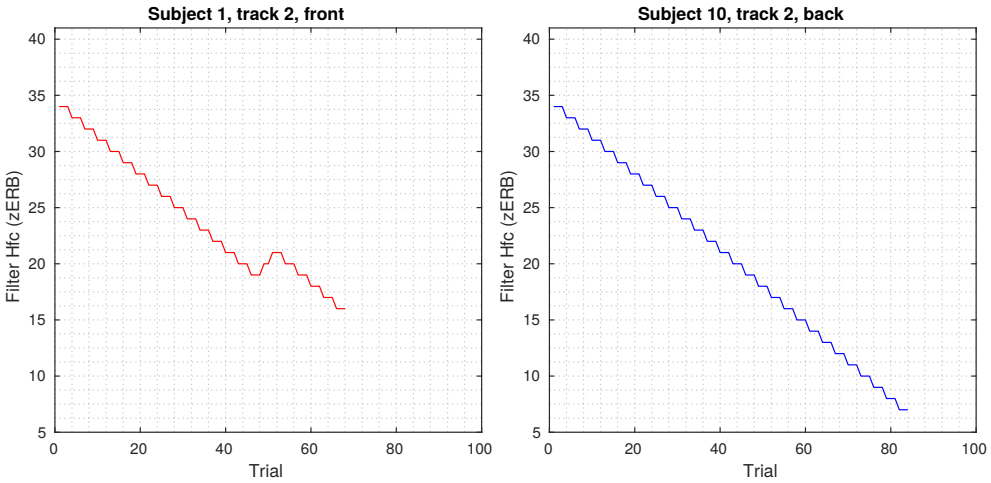


Figure 41: Examples of tracks displaying excellent localization performance despite narrow stimulus bandwidth.

A third group of subjects displayed inconsistent localization performance, suggesting that some subjects switched between the cues they attended to throughout the experiment. This is indicated by back and forth rowing of the adaptive tracks and sudden changes in localization performance. Examples of such tracks are plotted in Fig. 42. Referring for example to the plot at bottom left of said figure, it can be seen that track 2 of subject number 8, obtained from the rear hemiplane test consists of several distinct up-down runs of the adaptive track. These runs show fairly consistent trends in the direction of stimulus bandwidth change and span a total range of eight units on the ERB-rate scale corresponding to a bandwidth variation of over an octave. The wide range of the runs makes it unlikely that the reversals in the track represent any meaningful threshold for cue dominance, rather it is more likely that the runs are a result of cue switching. Similar observations suggestive of cue switching can be made about the other tracks displayed in Fig. 42.

7.4 Comparison of localization performance between the two hemiplanes

Comparison of the overall performance between the frontal and rear hemiplane experiments shows that the majority of subjects reached narrower stimulus bandwidths in the rear hemiplane experiment. This is indicated by the increased number of tracks that ended by crossing their lower limits in the rear hemiplane test. Table 4 shows that only two tracks terminated this way in the case of the frontal plane experiment whereas there were a total of thirteen such tracks in the rear hemiplane experiment. Similarly, the number of tracks that terminated by crossing the upper limit decreased from eleven to four between the frontal and rear hemiplane experiments.

One subject in particular performed very differently between the two experiments, as shown by Fig. 43. This subject finished all four tracks in the frontal plane experiment by crossing the upper limit of the tracks, i.e., the presented stimuli were systematically not localized to the front cluster. Conversely, the same subject was

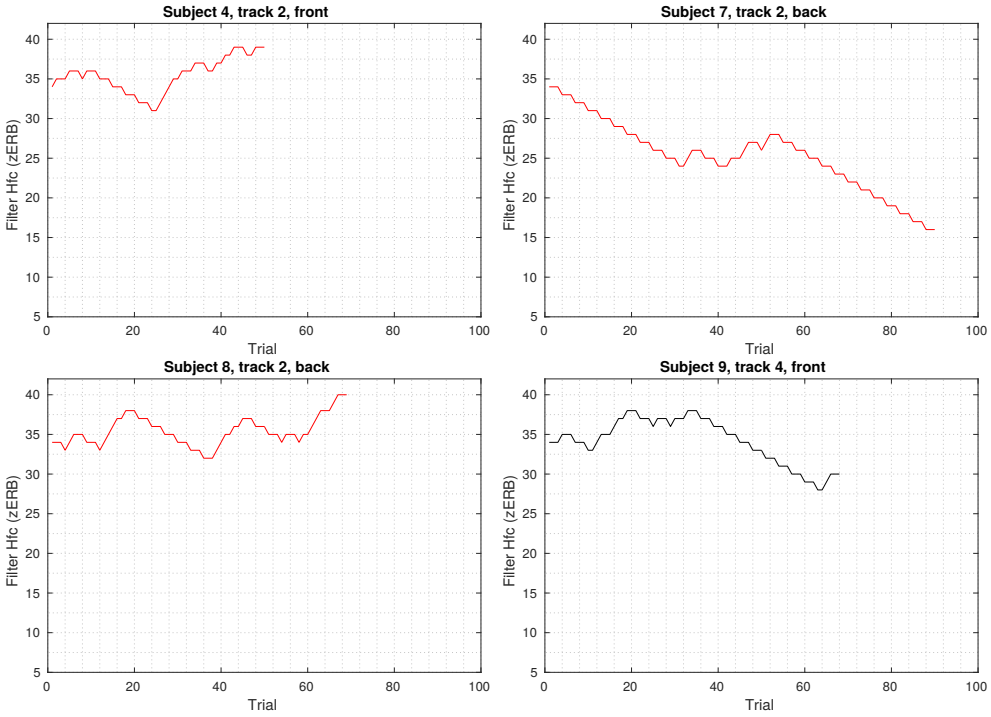


Figure 42: Examples of tracks showing inconsistent localization performance.

able to complete the test by crossing the lower limits of three of the four tracks in the rear hemiplane experiment. The results of this subject are similar to those reported by Kawaura et al. for one subject who experienced consistent front to rear reversals in the cue conflict experiment described in [18]. It is possible that other subjects experienced similar reversals in a less consistent manner. Such front to rear reversals could partly account for the increased number of tracks that terminated by crossing the lower limits in the rear hemiplane experiment.

7.5 Methodological considerations

There are several factors that should be taken into account in interpreting the results from the conducted experiment and considered in the design of any possible follow up experiment. This section outlines some of these issues and attempts to offer possible solutions.

7.5.1 Answering scheme

In both experiments, burst sequences that were not presented from the tested direction were dummy trials and had no effect on the adaptive tracks. It was therefore possible to complete the test by crossing the lower limits of all four tracks by systematically answering that the burst sequence was perceived in the tested direction. Although none of the subjects were aware of the underlying logic of the test procedure, it is possible that some subjects experienced localisation biases that went unnoticed by the test logic. This data would have been particularly useful in assessing the

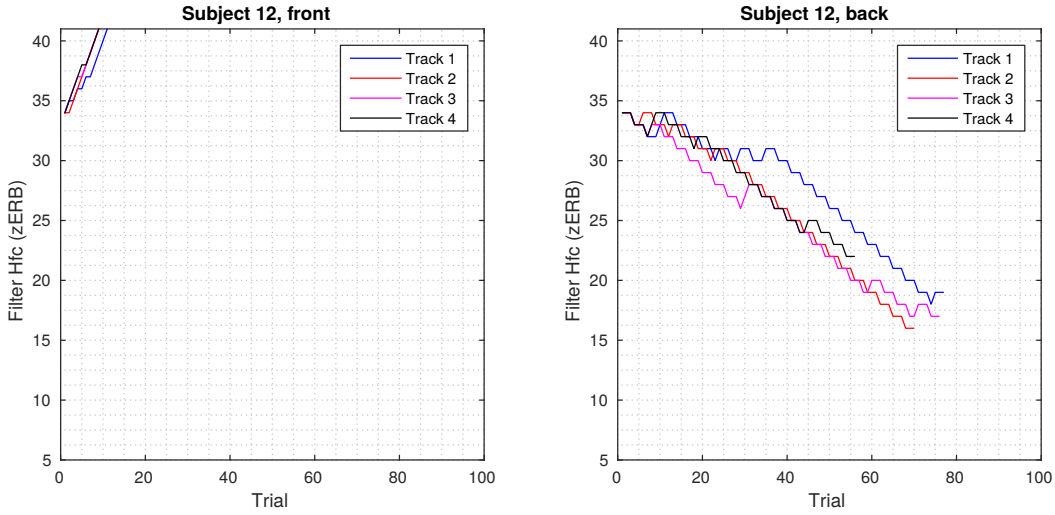


Figure 43: Results for subject 12. This subject displayed consistently poor localization performance in the frontal hemiplane experiment and consistently accurate performance in the rear hemiplane experiment. It is possible that this subject experienced systematic front to rear reversals under conflicting dynamic and spectral cues.

possibility of cue conflict resulting in systematic localization biases towards the rear hemiplane, that could account for the apparent increase in localization performance in the rear hemiplane experiment.

7.5.2 Duration of the experiment

Because the experiment followed an adaptive procedure, the number of trials each subject completed in a single session varied widely. Subject 12 for example completed only 92 trials in the frontal hemiplane experiment whereas subject 1 completed a total of 668 trials in the same experiment. It could be that the prolonged length of the experiment had a negative effect on the localization performance of some subjects. Fig. 44 shows two examples of two tracks that show a decline in localization performance as the test progressed. Both subjects localized the stimuli accurately in the beginning of the experiment, but showed a sudden change in performance after having completed a considerable number of trials. Note that the trial number shown on the x-axis denotes the trial number for the plotted track only; the total number of completed trials at the reversal point is a combination of the trials for all four tracks and the dummy trials.

The effective length of the experiment could be made shorted by reducing the number of dummy trials or adapting an entirely different test arrangement. One possibility is to use the method of constant stimuli that ensures that all subjects complete the same number of trials.

7.5.3 The effect of narrowband stimuli

Although all of the four tracks corresponded to wideband stimuli at the start of the test, good localization performance could reduce stimulus bandwidth well below an

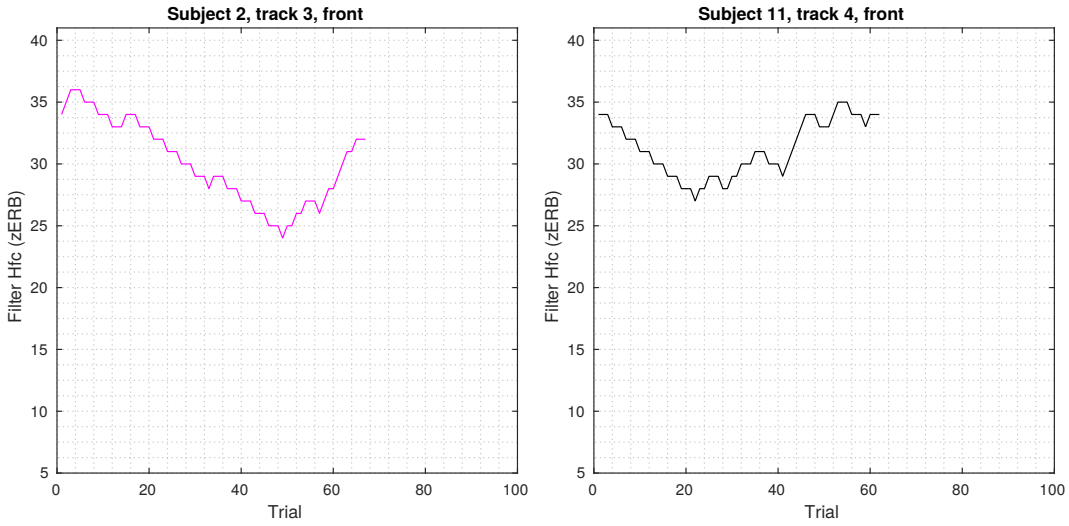


Figure 44: Tracks displaying reduced localization performance as the test progressed.

octave for some tracks. Blauert [2], [3] has shown that narrowband stimuli presented from a location coinciding with the median plane are localized based on their center frequency rather than the actual angle of incidence. Moreover, the results reported by Itoh et al. [16] indicate that the apparent source directions produced by this effect vary widely between subjects.

This effect could account for the diagonal source elevations reported by some subjects in the post-experiment discussions summarized previously in Section 7.2. It also poses a problem for the design of experiments where only two localization processes are assumed to be active in forming the localization judgement. In the context of the experiment described in this work for example, it violates the assumption that incorrectly localized burst sequences are a result of the misleading dynamic cues dominating the spectral cues as it is possible that the subjects answered incorrectly because they attended to the inaccurate spectral cues carried by narrowband stimuli, rather than the inaccurate dynamic cues formed by the head-motion-coupled source movement.

7.5.4 The effect of stimulus duration

Because the experimental arrangement made use of real sound sources exclusively, rather than virtual sources created via panning between two loudspeakers, the possible locations of stimulus presentation were limited by the angular separation of the loudspeakers. Once a burst stimulus had been triggered by the head tracker, the stimulus remained static throughout its 10 ms duration regardless of the instantaneous orientation of the subjects head. As discussed previously in Section 7.3, the fact that some subjects were able to localize narrowband burst sequences to the correct hemiplane seems to indicate that some subjects were able to extract accurate dynamic cues from the individual bursts in the stimulus sequence thus violating the fundamental assumptions of the experiment.

8 Summary and future work

This work investigated auditory localization under conflicting spectral and dynamic cues. The motivation for the work was laid out in Chapter 1 and the relevant theoretical background was presented in Chapters 2-4. Chapter 5 provided an overview of previous studies assessing the same topic.

In the experimental part of the thesis work described in Chapter 6, a listening test was designed and implemented in an anechoic chamber equipped with a multichannel loudspeaker system and motion-tracking cameras. The motion tracking cameras were used to move acoustic stimuli within the test chamber to produce misleading dynamic localization cues. The relative weighting of spectral and dynamic cues was investigated as a function of stimulus spectrum. Statistical analysis of the obtained data did not reveal a statistically significant correlation between the amount of low-frequency energy carried by the stimulus and the minimum stimulus high cutoff frequency required for accurate localization performance.

Overall, the obtained results appear to be in general agreement with previous studies in that no frequency threshold seems to exist above which spectral cues systematically dominate localization in the presence of conflicting dynamic cues. As a result, it appears that subjects can make localization judgements based on either of the cues provided by the stimulus, leading to inconsistent localization behaviour as the subjects alternate between the cue they attend to. The up-down rowing observed in several of the obtained adaptive tracks are suggestive of such localization behaviour. Some subjects appeared to favour one cue over the other, as displayed by the set of adaptive tracks that systematically approached either their upper or lower limits. Some subjects were able to localize narrowband stimuli deprived of accurate spectral cues correctly to the right hemiplane, possibly by extracting accurate dynamic cues from individual noise bursts in the stimulus sequences.

Localization performance was apparently better in the rear hemiplane as indicated by the increased number of adaptive tracks that terminated by crossing their lower limits. However, the results obtained for one subject are reminiscent of those reported by Itoh et al. in [16], where a single subject systematically localized stimuli with conflicting dynamic and spectral cues to the rear hemiplane in virtual auditory space. The apparent increase in localization accuracy in the rear hemiplane experiment could be accounted for by the subjects experiencing similar localization biases that were not captured by the test logic.

Informal discussions with the test subjects revealed that some subjects experienced localization phenomena that do not correspond to those predicted by accurate spectral cues or the principle of least displacement. Several subjects felt that the burst sequences were not localizable in any one direction, but rather experienced the burst sequences as diffuse. The head-motion-coupled panning arrangement reportedly led some subjects to internalize the free-field stimuli rather than localize the sequences to any external location.

In summary, despite the limitations of the experimental arrangement, the obtained results seem to indicate, that no threshold exists for the dominance of spectral cues over dynamic cues. Rather, the localization processes seem to function in parallel,

enabling localization judgements to be made based on either one of the available cues. Cue preference and rate of cue switching seems to vary widely between subjects and over time. Some subjects may localize stimuli systematically to the rear hemiplane when presented with conflicting spectral and dynamic cues. Using a displacement factor of one in a head-motion-coupled panning arrangement can lead some subjects to internalize acoustic stimuli presented over loudspeakers; similarly, some subjects experience a diffuse stimulus location not localizable to any one direction.

8.1 Future work

The experimental evidence reported in currently existing literature and the experimental results described in this work suggest, that rather than one cue showing a systematic tendency to dominate the other in localization judgements, subjects tend to show inconsistent localization performance in two-alternative forced choice tasks aimed at identifying the dominant cue. It therefore seems likely, that the localization processes function in parallel suggesting that the inconsistent localization behaviour observed in past studies is a result of the experimental arrangement where subjects were given a binary choice between the two localization percepts corresponding with the two processes. However, lateralization studies assessing the perception of acoustic stimuli providing conflicting ITD and ILD cues have been shown to result in the formation of two separate auditory images in some subjects, (see e.g., [15], [43]). An interesting direction for future work would be to investigate, whether this type of splitting of the auditory image can occur under conflicting spectral and dynamic cues.

One possibility for implementing such an experiment is to use a displacement factor of two to move a continuous noise stimulus in virtual auditory space resulting in a static phantom source in the hemiplane opposing that of stimulus presentation. In the experimental scenario, the subject would perform slow oscillating head rotations about a relatively small spatial window of five to ten degrees while a fixed length stimulus of several seconds was processed with the interaural delays and HRTFs corresponding to the displacement factor of two head coupled panning arrangement at any given instant. The task of the subject would be to report whether or not he/she perceives sound sources in both hemiplanes or not. The probability of the subjects perceiving a divided auditory image can be measured as a function of stimulus spectrum. Due to the fact that dynamic cues are most effectively extracted from low-frequency ITD cues, the rate of auditory image splitting should decrease when low-frequency components are removed from the stimulus; similar reduction in splitting rate is expected for low-pass filtered stimuli lacking high-frequency components.

Using a displacement factor of two should resolve the problem of directional bands associated with narrowband stimuli as the stimuli are no longer confined to the median plane of the subject. Likewise, using continuous stimuli presented from continuously updated coordinates in virtual auditory space rather than burst sequences over loudspeakers removes the possibility of the subjects extracting accurate dynamic cues from individual bursts presented from static locations.

References

[1] S. Bech and N. Zacharov. *Perceptual audio evaluation - Theory, Method and Application*. John Wiley & Sons, 2007.

[2] J. Blauert. Sound localization in the median plane. *Acta Acustica united with Acustica*, 22(4):205–213, 1969.

[3] J. Blauert. *Spatial Hearing: the Psychophysics of Human Sound Localization*. MIT press, 1997.

[4] W. O. Brimijoin and M. A. Akeroyd. The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, 3(3):179, 2012.

[5] W. O Brimijoin, A. W. Boyd, and M. A. Akeroyd. The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE*, 8, 2013.

[6] A. C. Burdette. *Analytic Geometry*. Academic Press, 2014.

[7] C. M. Douglas. *Design and Analysis of Experiments*. John Wiley & Sons, 2012.

[8] H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. Springer Science & Business Media, 2007.

[9] W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress. Localization of high-frequency tones. *The Journal of the Acoustical Society of America*, 29(9):988–991, 1957.

[10] H. Fletcher. Auditory patterns. *Reviews of Modern Physics*, 12(1):47, 1940.

[11] S. A. Gelfand. *Hearing: An Introduction to Psychological and Physiological Acoustics*. CRC Press, 2009.

[12] A. Ghasemi and S. Zahediasl. Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2):486, 2012.

[13] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1):103–138, 1990.

[14] G. V. Glass, P. D. Peckham, and J. R. Sanders. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, pages 237–288, 1972.

[15] E. R. Hafter and L. A. Jeffress. Two-image lateralization of tones and clicks. *The Journal of the Acoustical Society of America*, 44(2):563–569, 1968.

[16] M. Itoh, K. Iida, and M. Morimoto. Individual differences in directional bands in median plane localization. *Applied Acoustics*, 68(8):909–915, 2007.

[17] M. Karjalainen. *Kommunikaatioakustiikka*. Otamedia Oy, 2000.

[18] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone. Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *Journal of the Acoustical Society of Japan (E)*, 12(5):203–216, 1991.

[19] J. Kim, M. Barnett-Cowan, and E. A. Macpherson. Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues. In *Proceedings of Meetings on Acoustics*, volume 19. Acoustical Society of America, 2013.

[20] G. F. Kuhn. Physical acoustics and measurements pertaining to directional hearing. In *Directional Hearing*, pages 3–25. Springer, 1987.

[21] H. Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477, 1971.

[22] L. M. Lix, J. C. Keselman, and H. J. Keselman. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test. *Review of Educational Research*, 66(4):579–619, 1996.

[23] E. A. Macpherson. Head motion, spectral cues, and wallach's 'principle of least displacement'in sound localization. *Principles and Applications of Spatial Hearing*, pages 103–120, 2011.

[24] E. A. Macpherson. Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation. In *Proceedings of Meetings on Acoustics*, volume 19. Acoustical Society of America, 2013.

[25] J. C. Makous and J. C. Middlebrooks. Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, 87(5):2188–2200, 1990.

[26] W. Martens, S. Sakamoto, L. Miranda, and D. Cabrera. Dominance of head-motion-coupled directional cues over other cues during walking depends upon source spectrum. In *Proceedings of Meetings on Acoustics*, volume 19. Acoustical Society of America, 2013.

[27] W. L. Martens, D. Cabrera, and S. Kim. The 'phantom walker'illusion: Evidence for the dominance of dynamic interaural over spectral directional cues during walking. *Principles and applications of spatial hearing*, pages 159–158, 2011.

[28] K. I. McAnally and R. L. Martin. Sound localisation during illusory self-rotation. *Experimental Brain Research*, 185(2):337–340, 2008.

[29] J. C. Middlebrooks and D. M. Green. Sound localization by human listeners. *Annual Review of Psychology*, 42(1):135–159, 1991.

[30] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Brill, 2012.

[31] D. Morikawa and T. Hirahara. Signal frequency range necessary for horizontal sound localization. *Acoustical Science and Technology*, 31(6):417–419, 2010.

[32] R. D Patterson. Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, 59(3):640–654, 1976.

[33] S. Perrett and W. Noble. The contribution of head motion cues to localization of low-pass noise. *Perception & Psychophysics*, 59(7):1018–1026, 1997.

[34] S. Perrett and W. Noble. The effect of head rotations on vertical plane sound localization. *The Journal of the Acoustical Society of America*, 102(4):2325–2332, 1997.

[35] C. J. Plack. *The Sense of Hearing*. Psychology Press, 2013.

[36] V. Pulkki and M. Karjalainen. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. John Wiley & Sons, 2014.

[37] T. D. Rossing. *Springer Handbook of Acoustics*. Springer Science & Business Media, 2007.

[38] J. Schnupp, I. Nelken, and A. King. *Auditory neuroscience: Making Sense of Sound*. MIT Press, 2011.

[39] J.W Strutt. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232, 1907.

[40] W. R. Thurlow and P. S. Runge. Effect of induced head movements on localization of direction of sounds. *The Journal of the Acoustical Society of America*, 42(2):480–488, 1967.

[41] H. Wallach. On sound localization. *The Journal of the Acoustical Society of America*, 10(4):270–274, 1939.

[42] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339, 1940.

[43] R. H. Whitworth and L. A. Jeffress. Time vs intensity in the localization of tones. *The Journal of the Acoustical Society of America*, 33(7):925–929, 1961.

[44] F. L. Wightman and D. J. Kistler. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5):2841–2853, 1999.

[45] W. A. Yost and S. Sheft. Auditory perception. In *Human Psychophysics*, pages 193–236. Springer, 1993.

[46] P. T. Young. Auditory localization with acoustical transposition of the ears. *Journal of Experimental Psychology*, 11(6):399, 1928.

[47] P. T. Young. The rôle of head movements in auditory localization. *Journal of Experimental Psychology*, 14(2):95, 1931.